IDENTIFYING DRUG THERAPY INAPPROPRIATENESS: DETERMINING THE VALIDITY OF DRUG USE REVIEW SCREENING CRITERIA

by Ilene H. Zuckerman[1], Principal Investigator
Diane L. McNally[1], Project Director,
Frank J. Hooper[2], Stuart Speedie[3], Colleen J. Metge[4], David A. Knapp[1]

Current affiliations:
[1]Center on Drugs and Public Policy, School of Pharmacy, University of Maryland at Baltimore
[2]School of Medicine, University of Maryland at Baltimore
[3]School of Medicine, University of Minnesota
[4]Faculty of Pharmacy, University of Manitoba

Federal Project Officer: Kathleen Gondek

Center on Drugs and Public Policy,
School of Pharmacy
University of Maryland at Baltimore

Health Care Financing Administration Cooperative Agreement No. 18-C-90302

January 1997

JJ ZUCKERMA@PHARMACY.AB.UMD.EDU

ACKNOWLEDGMENTS

We thank Lori Walker for her programming assistance and Grace Roscoe for secretarial support. We recognize Penny Bard, Dr. Julie Kreyenbuhl, Dr. Leanne Lai, Vithaya Kulsomboon and Maneerat Rattanamahattana for their assistance in data collection and data entry. We are extremely grateful to Janet Freeze and her staff at the state of Maryland Department of Health and Mental Hygiene, Medical Care Finance and Compliance Administration, who helped us in obtaining Maryland Medicaid claims data. We also acknowledge the participation of the hypertension experts in the Delphi survey; we have listed their names in the text. In addition, we are grateful for the participation of our expert panel reviewers (Drs. Dobbin Chow, Robert Feroli, Michael Freedman, Joseph Ober, Mona Tsoukleris and Debra Wertheimer). Finally we thank the support and assistance of our project officer, Dr. Kathleen Gondek.

CONTENTS

List of Tables

Symbols

...	Category not applicable
---	Not statistically significant ($p>0.05$)
d.f.	degrees of freedom
S.D.	Standard Deviation
S.E.	Standard Error of the mean

EXECUTIVE SUMMARY

Background

The use of drug use screening criteria for application in outpatient Medicaid prescription drug programs is mandated by the Omnibus Budget Reconciliation Act of 1990. These criteria are used to screen prescription drug claims for their prescribing and dispensing inappropriateness.

Ultimately, validation of the use of DUR screening criteria to identify and intervene upon inappropriate drug therapy and prescribing will require outcome studies. Meanwhile, some intermediate measures of the usefulness of DUR screening criteria will be helpful to DUR Boards that must deal with the issue of outpatient DUR. While the ultimate goal of this project was to strengthen the ability of outpatient (DUR) screening criteria to identify clinically significant cases of inappropriate drug therapy, we focused on the aim of evaluating the validity of DUR computer-based screening using claims data. We selected treatment of hypertension as a suitable context for evaluation because: (a) hypertension is a prevalent disease in the general population, and is the most prevalent disease in our cohort; and (b) practice guidelines have been adopted and widely accepted for the diagnosis and treatment of hypertension. To accomplish this we set three specific objectives:

- Quantify the agreement between an outpatient drug use review screening of Medicaid claims data (DURSCREEN) assessment and a more in-depth review, clinical expert (INDEPTH) assessment of identifying drug therapy inappropriateness (construct validity).

- Test the hypothesis that subjects with appropriate antihypertensive drug therapy (as identified by drug use review screening) have lower mean blood pressures than subjects with inappropriate antihypertensive drug therapy (criterion validity).

- Produce a manual for drug use review programs across the country on how to assemble a minimal data set to permit an ongoing assessment of drug use review screening of Medicaid claims data when applied to other drugs, diseases and populations.

Every state is mandated by federal law to operate an outpatient drug use review program for Medicaid, *"to improve the quality of pharmaceutical care by ensuring that prescriptions are appropriate, medically necessary and that they are not likely to result in adverse medical events*" (Omnibus Budget Reconciliation Act of 1990). The overall intent of these programs is to employ validated criteria and a screening process to gauge the extent of "appropriateness" of drug therapy and to intervene on "inappropriateness" when subsequently identified. Study of the appropriateness of medical interventions is

not new given the last decade's spiraling health care costs and the need to be efficient in allocating scarce resources for medical care. The most common of medical interventions is drug therapy and attention to its "appropriateness" has escalated as demonstrated by Congress' action in October 1990. Although drug use review programs have expanded to include many potential problem areas, there has been a recent call to look at how "inappropriateness" is affecting quality of care and patient well-being (Lipton and Bird, 1993; Soumerai and Lipton, 1995).

This proposal examined the validity of the measurement of drug therapy inappropriateness, within the context of federally mandated (Omnibus Budget Reconciliation Act of 1990) drug use review programs. States are currently allocating substantial resources to carry out the Omnibus Budget Reconciliation Act of 1990 outpatient drug use review requirements. Under these requirements, states establish standards, identify patterns of inappropriate drug therapy and design and implement interventions to improve drug therapy appropriateness. However, we do not know if the drug therapy inappropriateness model (i.e., the drug use review model mandated by the Omnibus Budget Reconciliation Act of 1990) is valid: do estimated rates of drug therapy inappropriateness reported by outpatient drug use review screening correlate with true rates of inappropriateness?

Drug use review programs approach their task of identifying drug therapy appropriateness or inappropriateness by: (1) defining the scope of drug therapy to be reviewed (by identifying the frequency and costs of drug therapy use in the population); (2) convening an expert panel (in the law it is the state's drug use review Board) to review prescribed compendia and the peer-reviewed literature to develop screening criteria for indicating drug therapy inappropriateness; (3) using the drug use review Board to approve drug therapy inappropriateness criteria and to set acceptable standards for variation from the criteria; (4) applying the criteria to Medicaid claims data (i.e., a secondary data set)--drug use review screening--and assigning a nominal ranking (YES vs. NO) to indicate overall drug therapy inappropriateness; and (5) assessing patient outcomes from drug therapy.

There are several inherent problems in using drug use review for identifying drug therapy inappropriateness including: (1) its approach to identifying drug therapy inappropriateness from a perspective limited to drug therapy only rather than a global disease management perspective; (2) the current limitations in the database used to identify drug therapy inappropriateness; i.e., secondary data with outpatient disease codes when available (intended for billing purposes) are sometimes used as a source of diagnostic information rather than primary patient data; (3) the inability to detect "inappropriateness" when a drug therapy has not been started (i.e., the patient has a disease but not an indicated drug therapy); (4) use of criteria based on randomized *controlled* trial data but reflecting only the consensus of expert opinion on the effectiveness of drug therapy in an *uncontrolled* clinical environment; and (5) setting

standards and basing decisions using estimated rates of inappropriate drug therapy that may or may not reflect the true rates. These problems lead to classifying an episode of drug therapy as inappropriate when it is appropriate (false positive results), and classifying truly inappropriate episodes of drug therapy as appropriate (false negative results). A perfect test of drug therapy inappropriateness would occur when there are no false positives and no false negatives and the estimated rate of drug therapy inappropriateness equals the true rate of drug therapy inappropriateness. However, the true rate of drug therapy inappropriateness is difficult to measure in an outpatient clinical practice environment.

The requirements of the Omnibus Budget Reconciliation Act of 1990, however, are currently in operation. The Health Care Financing Administration, as the watchdog responsible for making sure that states meet their Federal financial participation requirements, ensures that the Federal financial participation-required outpatient drug use review programs are implemented. Consequently, states are currently allocating substantial resources to ensure drug therapy appropriateness.

The state-administered Medicaid outpatient drug use review programs are required to categorize drug therapy inappropriateness criteria using specific, predetermined elements: that is, whether there is therapeutic duplication (another drug that is being used for the same indication without additional benefit), drug-disease and drug-allergy contraindications, adverse drug-drug interactions, correct dose and duration of therapy, clinical abuse and misuse. And, states are required to set standards for acceptable variation from the screening criteria, identify patterns of inappropriate drug therapy and from these results, design and implement interventions to improve drug therapy appropriateness. Without knowledge of the sensitivity and specificity of drug use review screening programs (i.e., false positive and false negative results), drawing valid conclusions from drug use review program results and instituting cost-effective interventions for improving drug therapy appropriateness is difficult.

Drug use review under the Omnibus Budget Reconciliation Act of 1990 mandated model may be an efficient method of screening for drug therapy inappropriateness. Claims data allow one to quickly screen many prescriptions with computer-applied algorithms of drug use criteria to identify drug therapy inappropriateness. We do not know however, if this efficient method is valid. In other words, do estimated rates for drug therapy inappropriateness from drug use review screening of claims data correlate with true rates of inappropriateness? Given that we do not know this true rate or even its approximation (there is an absence of a "gold standard"), how then do we establish standards for acceptable variation from the drug use criteria? The current state of federally mandated drug use review is that we do not know whether drug use review screening of claims data validly measures drug therapy inappropriateness; effectiveness studies to show this are not yet available. Such studies are very costly and labor-intensive because they require access to and collection of primary data. Effectiveness

studies also require a strongly correlated measure of patient outcomes with drug therapy, e.g., antihypertensive therapy to control of blood pressure and subsequent prevention of strokes.

Faced with uncertainty in the ability of this currently mandated program to accurately measure drug therapy inappropriateness, we examined the specificity and sensitivity of the current assessment of determining drug therapy inappropriateness. We did this by applying drug use review screening criteria to Medicaid claims data (DURSCREEN) and comparing this assessment of drug therapy inappropriateness with a clinical expert (INDEPTH) assessment of identifying drug therapy inappropriateness.

Methods

Medicaid patients with evidence of a diagnosis of hypertension were identified from primary medical record data abstraction. Secondary administrative claims data on these patients were available on tape from the state of Maryland's Medicaid program.

We chose hypertension as a disease to study the validity of a test of medication inappropriateness because: (a) hypertension is a prevalent disease in the general population, and is the most prevalent disease in our cohort; and (b) practice guidelines have been adopted and widely accepted for the diagnosis and treatment of hypertension.

First, a panel of experts in hypertension was asked to agree on explicit criteria for judging hypertensive drug therapy inappropriateness. Agreement was accomplished using a Delphi technique. Second, three physician/pharmacist clinician pairs were recruited and trained to apply the explicit drug use review criteria to a subject profile containing primary clinical data and secondary claims data. Using a structured implicit review process, the reviewers assessed the subject's antihypertensive drug therapy as appropriate, inappropriate or indeterminate. A consensus process was developed to adjudicate disagreements between paired clinicians. This process of clinician review of subject profiles containing primary and secondary data was termed the INDEPTH assessment.

Simultaneously, the explicit drug use review screening criteria were applied to a data set of our cohort's Medicaid service and prescription claims to screen for drug therapy inappropriateness. This involved developing computerized algorithms that applied the intent of the criteria as accurately as possible. We called this the DURSCREEN assessment. The DURSCREEN assessment used secondary claims data only.

Third, the results of the DURSCREEN assessment were compared with identification of drug therapy inappropriateness found from the INDEPTH assessment. Fourth, the relative "specificity and sensitivity" of specific DURSCREEN criteria elements (singly and in combination) were determined from the INDEPTH assessment. We employed

contingency table analysis, construction of receiver operating characteristic (ROC) curves and logistic regression techniques.

Fifth, mean blood pressures of subjects with appropriate drug therapy as identified by DURSCREEN were compared with mean blood pressures of subjects with inappropriate drug therapy to test the hypothesis that subjects with appropriate antihypertensive drug therapy have lower mean blood pressures than subjects with inappropriate antihypertensive drug therapy.

Lastly, a manual of operations was developed that explains how to assemble a minimal data set to permit an ongoing assessment of drug use review screening of Medicaid claims data when applied to other drugs, diseases and populations.

Summary of Results

The INDEPTH assessment of antihypertensive drug therapy inappropriateness was designed to approximate a "gold standard" measure. To facilitate this clinical expert INDEPTH assessment of drug therapy inappropriateness, Medicaid hypertensive patient profiles were built from several information sources. Of the original 788 subjects identified from primary medical data record abstraction, 738 were eligible for analysis. One hundred of these subjects were labeled as "indeterminate" when they could not be classified as having either appropriate or inappropriate antihypertensive drug therapy using the INDEPTH assessment. Of the remaining 638 study subjects, nearly 25% were identified as having inappropriate drug therapy.

The demographic profiles of subjects with appropriate and inappropriate drug therapy were compared with the group of 100 indeterminate study subjects. No differences were found for sex, race, age, and the number of disease categories. Seventeen percent of indeterminate subjects did not have a single blood pressure reading. Of the remaining 83 subjects, more than 90% had indications of "uncontrolled" blood pressure but scant data were available to follow the course of therapy during the brief period used for assessment. The distinguishing feature for panelists "labeling" these subjects as having "indeterminate" appropriateness was missing data.

The main validation feature for the INDEPTH assessment focused on blood pressure control. The group of study subjects with appropriate antihypertensive drug therapy consistently demonstrated significantly lower blood pressure readings than the group of study subjects with inappropriate antihypertensive drug therapy. The percent of "uncontrolled" blood pressure readings was shown to be significantly higher among the group of subjects identified with inappropriate drug therapy. These findings provide evidence for the validity of INDEPTH assessment as a measure of antihypertensive drug therapy inappropriateness.

Fifty-three distinctive computer-based decision algorithms were used to translate the 92 drug use screening criteria resulting from the Delphi survey for use in the DURSCREEN assessment. These algorithms were then used to identify drug therapy inappropriateness for the 738 study subjects using administrative data from Medicaid claims. A single instance of any criterion exception, or flag, was considered inappropriate therapy. Nearly two-thirds of all study subjects were identified as inappropriate. A total of 201 (43%) subjects classified by DURSCREEN as "inappropriate" failed more than one criterion. Utilization (both over- and under-utilization) was the primary identifier for drug therapy inappropriateness. The number of DURSCREEN flags per subject ranged from zero to ten; the mode was zero and the median was one flag. The median number of criteria elements (i.e., dose, duplication, drug-drug interaction, drug-disease contraindication, over-utilization, under-utilization) failed per subject was one.

The comparison of the basic screening instrument, DURSCREEN, with the INDEPTH assessment findings demonstrated statistically significant associations but very poor agreement (48%). The measure of sensitivity was 0.735 compared to a much lower specificity (0.395). Nine alternative DURSCREEN derivatives demonstrated varying levels of agreement, sensitivity, specificity and statistical association. These derivatives consisted of several combinations of the screening algorithms to operationally define drug therapy inappropriateness. One derivative, DURSCREEN(5) offered a middle of the ground approach with a 61.9% agreement rate, and measures of sensitivity and specificity of 0.561 and 0.638, respectively. DURSCREEN(5) defined inappropriateness as those subjects who failed at least one of the drug use screening criteria, but excluding subjects who failed *only* the under-utilization criterion.

Construction of receiver-operating-characteristic curves was used in an attempt to "improve" the statistical relationship between DURSCREEN and the INDEPTH findings. The number of DURSCREEN flags and the number of different criteria elements with flags were explored. Although all areas-under-the-curve were statistically different from chance occurrence, they were not clinically significant from 0.5 (range 0.6011-0.6568). The height and skewness of the curves provided little assistance in selecting a cutoff for maximum sensitivity and specificity of the DURSCREEN based on the number of flags.

A series of multivariate models was developed using two continuous measures of blood pressure (i.e., mean systolic blood pressure and mean diastolic blood pressure) as the dependent variable. The development of each model included a single measure of the computerized DURSCREEN (the original and one of nine derivatives) and four control variables identified as clinically and statistically important in model development (age, compliance ratio, the number of antihypertensive drugs prescribed and the number of disease categories). These models were used to test the hypothesis that inappropriate antihypertensive drug therapy (as identified by DURSCREEN) is associated with statistically significantly higher blood pressures than appropriate antihypertensive drug

therapy. Although many of the multivariate models were statistically predictive of blood pressure, no single DURSCREEN model emerged as the best model and the explanatory variance was low (range 3-10%). In only three models did the DURSCREEN measures show statistically significant *p*-values in their respective models. However, the explanatory variance was only 5%. Multiple regression on select DURSCREEN criteria and control variables demonstrated that individual DURSCREEN criteria did not provide statistical insight into the expressions of blood pressure assessment.

Limitations

We have identified several limitations to our findings. First, our study focused only on inappropriateness of antihypertensive drug therapy. Generalizing our findings to inappropriateness measures for treatment of other diseases and conditions would be premature. However, our study offers the framework for reproducing the methodology to use with other disease states.

Our cohort does not represent the general population. There is an over-representation of African Americans and the cohort is drawn from the Maryland Medicaid population treated in hospital-based clinics. Another important limitation is that this was a cross-sectional study, and was not designed to measure outcomes of inappropriate drug therapy. Because of the cross-sectional observational design, not all subjects were evaluated using the same amount of data. Although the period of observation was the same for each subject, data such as blood pressures and laboratory values were dependent on other factors. Specifically, the amount of data was determined by the number of primary care visits the subject experienced during the study period. We did not attempt to "weight" the value of subject data based on the number of clinic visits.

The cross-sectional design of our study limited our ability to collect additional variables that may have improved the predictability of our regression models. For example, many important variables (body mass index, diet, marital status) were not available for this study. Additionally, we used mean blood pressure measurements as our dependent variable, which did not allow us to examine the temporal relationship between changes in blood pressure and the presence of inappropriate drug therapy. Given these limitations, however, the models strongly suggested that individual drug use criteria are poorly related to blood pressure.

Our assumption that the INDEPTH assessment is the closest we have to a gold standard is a limitation, especially in the interpretation of the results of our study. One could argue that, although our INDEPTH assessment was statistically and clinically associated with effectiveness (blood pressure), we did not attempt to evaluate any association with adverse drug therapy outcomes. To demonstrate this relationship, a prospective, longitudinal study design should be employed, since adverse events are relatively rare. A prospective design would allow collection of necessary data (e.g.,

serum drug concentrations) to identify whether a drug-drug interaction resulted in an adverse event. A longitudinal study would give one the opportunity for a longer observational period to capture true incidence rates of clinically significant adverse drug therapy events. Despite these limitations, the INDEPTH assessment has utility as a measure of "truth" in the assessment of drug therapy inappropriateness.

Conclusions

We conclude that the drug use screening criteria in the form of computerized algorithms applied to administrative claims data are not sufficiently sensitive or specific in identifying patients with inappropriate antihypertensive drug therapy. Claims data are not rich enough to provide clinical insight into the subject's medical history. It appears that this clinical insight is a prerequisite for assessing drug prescribing offered by routine claims processing and monitoring of medical record databases.

We acknowledge that outpatient drug use review programs as mandated by federal legislation were intended to be a *screening* process for *potentially* inappropriate drug therapy. The screening, for drug use review, consists of applying content validated criteria to a patient's medication history. To allocate resources and run efficient drug use review programs, policy makers need to know the sensitivity and specificity of their drug use review programs. They should consider resources spent on false positive flags as well as the public health risk for false negative flags. This research yielded information on the sensitivity and specificity of a computerized drug use review screening program focusing on treatment of hypertension. Neither the sensitivity nor specificity was sufficiently high enough for an efficient screen. Improvement in the application of utilization criteria may improve the program's specificity. However, we conclude that a highly specific and sensitive screen requires more information than is currently available through administrative claims data. Specifically, clinical markers of drug therapy effectiveness may significantly improve the screen's sensitivity and specificity. Incorporation of clinical data should be feasible, especially for managed care organizations that take advantage of computerized medical records. Drug use review program managers should encourage development of this technology. In addition, programs that employ computerized algorithms for drug use review screening should use caution in denying payment or basing clinical decisions solely on such mechanisms.

A manual of operations has been developed to help those policy makers evaluating drug use review programs, whether in fee-for-service Medicaid programs or managed care environments. When selecting a drug use review vendor, one should assure that there has been an assessment of the program's validity; the manual offers a methodology to do so. Unless policy makers demand quality drug use review programs from vendors, the state of the art for drug use review will not improve, and resources will be wasted on ineffective, inefficient drug use review programs.

However, it would be costly and unrealistic to fully assess a DUR program's sensitivity and specificity. Alternatively, we recommend that prospective DUR screens be limited to those that, if violated, could lead to an immediate, identifiable threat to patient health. Use of other screening criteria should be limited to retrospective analyses examining drug prescribing patterns rather than identifying inappropriate drug therapy in individual patients.

INTRODUCTION

Objectives

The goal of this project was to strengthen the ability of outpatient drug use review (DUR) screening criteria to identify clinically significant cases of inappropriate drug therapy in the Medicaid program. The use of drug use screening criteria for application in outpatient Medicaid prescription drug programs is mandated by the Omnibus Budget Reconciliation Act of 1990 (Omnibus Budget Reconciliation Act, 1990). These criteria are used to screen prescription drug claims for prescribing and dispensing inappropriateness. At least one public-domain set of drug use screening criteria has been developed using a combination of the official compendia, approved labeling, the peer-reviewed literature, and expert panels of scientists and clinicians as a basis for widespread approval and acceptance (Knapp *et al.*, 1992). Screening criteria were developed for eight classes of drugs (angiotensin converting enzyme inhibitors, benzodiazepines, calcium channel blockers, cardiac glycosides, heterocyclic antidepressants, histamine H_2-receptor antagonists, non-steroidal anti-inflammatory drugs, psychotropics) and were designed to be applied to a minimal data set, such as that as available from a prescription claims database.

Ultimately, validation of the use of DUR screening criteria to identify and intervene upon inappropriate drug therapy and prescribing will require outcome studies. Meanwhile, some intermediate measures of the usefulness of DUR screening criteria will be helpful to DUR Boards that must deal with the issue of outpatient DUR. While the ultimate goal of this project was to strengthen the ability of outpatient (DUR) screening criteria to identify clinically significant cases of inappropriate drug therapy, we focused on the aim of evaluating the validity of DUR computer-based screening using claims data. We selected treatment of hypertension as a suitable context for evaluation because: (a) hypertension is a prevalent disease in the general population, and is the most prevalent disease in our cohort; and (b) practice guidelines have been adopted and widely accepted for the diagnosis and treatment of hypertension. To accomplish this we set three specific objectives:

- Quantify the agreement between an outpatient DUR screening of Medicaid claims data (DURSCREEN) assessment and a more in-depth review, clinical expert (INDEPTH) assessment of identifying drug therapy inappropriateness (construct validity).

- Test the hypothesis that subjects with appropriate antihypertensive drug therapy (as identified by DUR screening) have lower mean blood pressures than subjects with inappropriate antihypertensive drug therapy (criterion validity).

- Produce a manual for DUR programs across the country on how to assemble a minimal data set to permit an ongoing assessment of DUR screening of Medicaid claims data when applied to other drugs, diseases and populations.

Background and Importance

Every state is mandated by federal law to operate an outpatient drug use review (DUR) program for Medicaid, *"to improve the quality of pharmaceutical care by ensuring that prescriptions are appropriate, medically necessary and that they are not likely to result in adverse medical events"* (Omnibus Budget Reconciliation Act, 1990). Details of the law can be found in Section 1927 (g) of the Social Security Act as passed under the Omnibus Budget Reconciliation Act (OBRA) of 1990. The overall intent of these programs is to employ validated criteria and a screening process to gauge the extent of "appropriateness" of drug therapy and to intervene on "inappropriateness" when subsequently identified. Brook defines the provision of a particular intervention or service to a class of patients as "appropriate" when the benefits of providing the intervention exceed the risks associated with such care (Brook and McGlynn, 1991). Study of the appropriateness of medical interventions is not new given the last decade's spiraling health care costs and the need to be efficient in allocating scarce resources for medical care. The most common of medical interventions is drug therapy and attention to its "appropriateness" has escalated as demonstrated by Congress' action in October 1990 (OBRA).

Appropriateness of Drug Therapy

Like the findings in studies of appropriateness of medical interventions and procedures (Bernstein *et al.*, 1993(a); Bernstein *et al.*, 1993(b); Brook *et al.*, 1990; Chassin *et al.*, 1987; Gloor, Kissoon and Jourbert, 1993; Havens *et al.*, 1993; Hilborne *et al.*, 1993; Leape *et al.*, 1990; Leape *et al.*, 1993; Siu *et al.*, 1986; Winslow *et al.*, 1988), drug therapy has also been found to be suboptimal or "inappropriate" in both inpatient and outpatient settings review of drug use (Helling, Norwood and Donner, 1982; Laporte, Porta and Capella, 1983; Mas and Laporte, 1983; Stander and Yates, 1988; Wells, Goldberg and Brook, 1988). Estimated rates of inappropriate treatment have ranged from 15% to 30% for medical interventions and for outpatient drug therapy have ranged from less than 1% to 35% (Chrischilles *et al.*, 1988; Knapp *et al.*, 1992).

"Inappropriateness" for a medical intervention is usually indicated by a "yes or no," "should have treated or should not have treated." However, for drug therapy, "inappropriateness" is defined as the presence of at least one problem: inappropriate daily dose, inappropriate duplication of therapy, interacting drug combinations or inappropriate days supply (Chrischilles *et al.*, 1988; Mead and McGhan, 1988), that could increase the likelihood of an adverse medical event. For example, a study by Pouleur and colleagues suggested that insufficient daily dose or the inadequacy of administration, or

both, might be responsible for different degrees of angiotensin converting enzyme inhibition between groups taking enalapril and captopril and therefore for higher mortality in the captopril group (Pouleur *et al.*, 1991). Holt and coworkers found hospital admissions due to therapeutic overdosing with digoxin (Holt, Kundu and Forecast, 1978); this is direct evidence of inappropriateness. Cardiovascular drugs are not the only categories implicated in less than optimal patient outcomes from inappropriate drug therapy. Psychoactive medications have been implicated in adverse experiences of many kinds. Several studies have focused on their misuse or suboptimal use (Ray, Federspiel and Schaffner, 1980; Avorn *et al.*, 1989; Beers *et al.*, 1988), increased length of stay (Knapp *et al.*, 1980), increased risk of hip fracture (Ray *et al.*, 1987; Ray, Griffin and Downey, 1989), falls (Tinetti and Speechley, 1989; Granek *et al.*, 1987) and accidental injury (Oster *et al.*, 1990). Also, antipsychotic use has been associated with self reports of neuroendocrine adverse effects among women (Zito, Sofair and Jaeger, 1990). In addition, several investigators have examined appropriateness of prescribing which has been linked to appropriateness of drug therapy under current definitions of the process of DUR (Gurwitz, Soumerai and Avorn, 1990, Ingman *et al.*, 1975, Jones *et al.*, 1987, Knapp *et al.*, 1973, Knapp *et al.*, 1978, Palumbo *et al.*, 1978).

Drug Use Review

DUR is a structured and continuing program that reviews, analyzes, and interprets patterns and instances of drug use against predetermined criteria and standards. The concept of DUR had its origins in the late 1960's as part of the recommendations from the Task Force on Prescription Drugs (United States Department of Health, Education and Welfare, 1969). Then, DUR was described as a process that improves the quality of patient care both by reducing irrational prescribing and minimizing the consequent unnecessary expenditures. Since then, drug use review has evolved into a more dynamic process aimed at identifying not only inappropriate prescribing by the provider but also, inappropriate dispensing by the pharmacist and inappropriate use or consumption of a drug by the patient. DUR explicitly assumes that a positive patient outcome is most likely to occur when all participants in the drug use process engage in the most appropriate behavior. This "evolved" definition then, brought us back around to the commonly understood definition of "appropriateness" used earlier, that benefits (positive patient outcomes) exceed risks (negative patient outcomes). Although DUR programs have expanded to include many potential problem areas, there has been a recent call to look at how "inappropriateness" is affecting quality of care and patient well-being (Lipton and Bird, 1993; Soumerai and Lipton, 1995).

Prospective DUR, and subsequent identification of one or more problems (therapeutic duplication, drug-drug interactions, incorrect drug dosage, etc.) indicating drug therapy inappropriateness, has been suggested as a reason for denying payment for drug therapy in publicly paid programs such as Medicaid and the PACE Program in Pennsylvania. However, knowing the prevalence of a particular drug therapy problem and its effect on

patient outcomes is rare (Lipton and Bird, 1993). Lacking data on drug therapy problem prevalence and the effect of problems on patient outcomes through effectiveness studies, an intermediate evaluation of DUR should take place to ensure that administrative interventions that may restrict or limit drug prescribing are based on more than efficacy data. Without a comparative standard of effectiveness for DUR specific to patient, population, drug and disease state, a concentration on the reliability and validity of DUR as a method for detecting drug therapy inappropriateness is in order for this interim evaluation. Analysis of intra- and inter-rater reliability, the use receiver operating characteristic (ROC) analysis (Hanley and McNeil, 1982; Hanley and McNeil, 1983; Metz, 1978; Phelps, 1993) and the comparison of mean blood pressures, both systolic and diastolic, between those judged to be with and without appropriate drug therapy will help this study get as close to the "effectiveness" studies ultimately required for validation of the DUR process.

Measuring Inappropriateness of Drug Therapy

Review of the medical record has always been a means for evaluating the quality and appropriateness of inpatient care and it continues to be used to assess adherence to medical care standards despite its limitations (Donabedian, 1980; Donabedian, 1982; Dans, Weiner and Otier, 1985; Rubenstein *et al.*, 1990). In fact, implicit review by peers is generally considered the community standard for final quality decisions (Dans, Weiner and Otier, 1985; Rubenstein *et al.*, 1990). Structured implicit review is a process by which a reviewer's attention is serially focused on important aspects of care (Donabedian, 1982; Rubenstein *et al.*, 1990; Kahn *et al.*, 1989); reliability and validity of the review are improved by obtaining implicit judgments about the appropriateness of this care (Hayward, McMahon and Bernard, 1993).

Hanlon and colleagues reported on the intra- and inter-rater reliability of a method for assessing drug therapy appropriateness using an index of ten general criteria for medication appropriateness (Hanlon *et al.*, 1992). Overall inter-rater agreement for appropriateness was 0.88 and for inappropriateness was 0.95; the overall kappa was 0.83. Intra-rater agreement was as high with an overall kappa of 0.92. In a further study on the index's content validity, each of the 10 criteria for drug therapy appropriateness was weighted by a survey of eight physicians and two pharmacists to allow for a single summated score of appropriateness per medication. Putative heterogeneity of appropriateness scores was found as an indication of content validity when the summated score was compared to another population and reliability of the index remained high (intraclass correlation coefficient=0.74).

The structure of our INDEPTH assessment takes into account the findings from other appropriateness studies; that is, that reliability is enhanced by using a three-part scale to indicate appropriateness (i.e., the drug therapy as inappropriate, appropriate or equivocal)

(Hanlon *et al.*, 1992, Brook, 1989) and that validity follows in succession from a rigorous determination of content, construct and criterion-related validity.

Drug Therapy Inappropriateness and Hypertension

Elevated blood pressure for the most part is an asymptomatic disease whose long term effects usually become known only after an unusually long time (1992 Joint National Committee; Cruickshank, Thorp and Zacharias, 1987). Consequently it is often called the "silent disease" and its treatment challenge rests in convincing a patient to take drug therapy for a disease they "can't feel" (Elsen *et al.*, 1990; Psaty *et al.*, 1990; Col, Fanale and Kronholm, 1990). An important factor then, in the management of hypertension is the extent to which patients comply with their treatment regimen. It is also important that physicians support the current consensus recommendations for treatment (1992 Joint National Committee). Despite the difficulties faced in treating this disease, however, there have been improvements in blood-pressure control (Berkson 1980; Folsom 1983; MMWR 1990) which have in turn reduced the incidence of stroke and ischemic heart disease (Ostfeld 1990; Goldman 1984; Thom 1988). These same gains have unfortunately not been seen in minority populations who are poor, with lower educational levels. Although both access to care and patient noncompliance with their antihypertensive regimen have been found as predisposing factors for hypertensive emergencies and urgencies, physician "noncompliance" with guidelines for hypertensive treatment is yet to be studied in a thorough way.

This study examined the validity of measures of drug therapy inappropriateness, within the context of federally mandated (OBRA 1990) drug use review programs. States are currently allocating substantial resources to implement OBRA 1990 outpatient DUR requirements. Under these requirements, states establish criteria and standards, identify patterns of inappropriate drug therapy and design and implement interventions to improve drug therapy appropriateness. However, we do not know if the drug therapy inappropriateness model (i.e., the DUR model mandated by OBRA 1990) is valid: do estimated rates of drug therapy inappropriateness from outpatient DUR screening correlate with true rates of inappropriateness?

Researchers using measures of appropriate care have investigated how they are created but have done little study regarding the validity and application of the methods used to determine "appropriateness." A common approach used in identifying appropriateness includes (1) defining a medical intervention (e.g., surgery, drug therapy); (2) reviewing the literature to find indications that justify the intervention; (3) convening an expert panel to rank the indications on an appropriateness scale; (4) identifying a sample of patients with the indication and assigning an appropriateness score for each individual's intervention; and (5) assessing patient outcomes from an intervention.

DUR programs follow a similar method of identifying drug therapy inappropriateness. They approach their task by: (1) defining the scope of drug therapy to be reviewed (by identifying the frequency and costs of drug therapy use in the population); (2) convening an expert panel (in the law it is the state's DUR Board) to review prescribed compendia and the peer-reviewed literature to develop screening criteria for indicating drug therapy inappropriateness; (3) using the DUR Board to approve drug therapy inappropriateness criteria and to set acceptable standards for variation from the criteria; (4) applying the criteria to Medicaid claims data (i.e., a secondary data set)--DUR screening--and assigning a nominal ranking (YES vs. NO) to indicate overall drug therapy inappropriateness; and (5) assessing patient outcomes of drug therapy.

There are several inherent problems in using DUR for identifying drug therapy inappropriateness including: (1) its approach to identifying drug therapy inappropriateness from a narrow drug therapy perspective rather than a global disease management perspective; (2) the current limitations in the database used to identify drug therapy inappropriateness; i.e., secondary data with outpatient ICD-9-CM (International Classification of Diseases, 9th Revision, Clinical Modification) disease codes when available (intended for billing purposes) are sometimes used as a source of diagnostic information rather than primary patient data; (3) the inability to detect "inappropriateness" when a drug therapy has not been started (i.e., the patient has a disease but not an indicated drug therapy); (4) use of criteria based on randomized *controlled* trial data but reflecting only the consensus of expert opinion on the effectiveness of drug therapy in an *uncontrolled* clinical environment; and (5) setting standards and basing decisions using estimated rates of inappropriate drug therapy that may or may not reflect the true rates. These problems lead to classifying an episode of drug therapy as inappropriate when it was appropriate (false positive results); and classifying truly inappropriate episodes of drug therapy as appropriate (false negative results). A perfect test of drug therapy inappropriateness would occur when there are no false positives and no false negatives and the estimated rate of drug therapy inappropriateness equals the true rate of drug therapy inappropriateness. However, the true rate of drug therapy inappropriateness is difficult to measure in an outpatient clinical practice environment.

However, the requirements of OBRA 1990 are in operation now, and the Health Care Financing Administration (HCFA), as the watchdog responsible for making sure that states meet their Federal financial participation (FFP) requirements, ensures that the FFP-required outpatient DUR programs are implemented. Consequently, states are currently allocating substantial resources to ensure drug therapy appropriateness. The state-administered outpatient DUR programs are required to use the appropriateness of care model as illustrated above to identify drug therapy inappropriateness in the Medicaid population. In addition, they are required to categorize drug therapy inappropriateness criteria: that is, using specific, predetermined elements: whether there is therapeutic

duplication (another drug that is being used for the same indication), drug-disease and drug-allergy contraindication, adverse drug-drug interactions, correct dose and duration of therapy, clinical abuse and misuse. Finally, state governments are required to set standards for acceptable variation from the screening criteria, identify patterns of inappropriate drug therapy and from these results, design and implement interventions to improve drug therapy appropriateness. Without knowledge of the sensitivity and specificity of DUR screening programs (i.e., false positive and false negative results), drawing valid conclusions from DUR program results and implementing cost-effective interventions for improving drug therapy appropriateness is difficult. In other words, it would be a waste of substantial state and federal resources to develop and implement intervention programs based on incorrect information.

Importance of this Research

DUR under the OBRA 1990 mandated model may be an efficient method of screening for drug therapy inappropriateness. Claims data allow one to quickly screen many prescriptions with computer-applied algorithms of inappropriateness criteria to identify drug therapy inappropriateness. We do not know however, if this efficient method is valid: in other words, do estimated rates for drug therapy inappropriateness from DUR screening of claims data correlate with true rates of inappropriateness? The current state of federally mandated DUR is that we do not know whether DUR screening of claims data validly measures drug therapy inappropriateness; effectiveness studies to show this are not yet available[1]. Such studies are very costly and labor-intensive because they require access to and collection of primary data. Effectiveness studies also require a strong correlated measure of patient outcomes with drug therapy, e.g., antihypertensive therapy to control of blood pressure and subsequent prevention of strokes.

HCFA has funded a demonstration project on the effectiveness of prospective DUR in the context of the OBRA 1990 model of DUR. Prospective DUR is a review of a patient's drug therapy either before prescribing, dispensing or administering of the medication. Drawing any conclusions about the results of this demonstration may be difficult if we do not have some indication of the sensitivity and specificity of the DUR screening method used to identify drug therapy inappropriateness.

Faced with uncertainty in the ability of currently mandated programs to accurately measure drug therapy inappropriateness, we examined the specificity and sensitivity of the current assessment of determining drug therapy inappropriateness. We did this by applying drug use review screening criteria to Medicaid prescription drug claims data

[1]OBRA 1990 mandated demonstrations related to drug therapy interventions by pharmacists. The evaluation reports of the two funded drug use review demonstrations projects (Contract #500-93-0002) are due in 1998. The demonstrations and their evaluation are structured to give us some evidence of the "effectiveness" of DUR.

(DURSCREEN) and comparing this assessment of drug therapy inappropriateness with a clinical expert (INDEPTH) assessment of drug therapy inappropriateness. The results of this research may be used by state Medicaid program policy makers who are responsible for ensuring drug therapy appropriateness. HCFA should be especially interested in this study since HCFA is the federal agency responsible for ensuring that states comply with OBRA 1990 legislation; state agencies and individuals involved in implementing DUR programs under Medicaid frequently look to HCFA for guidance in designing DUR programs. In addition, HCFA is responsible for overseeing the evaluation of the prospective DUR demonstration project (HCFA Contract #500-93-0002), and recommending policy based on the results. It is imperative that HCFA policy makers know the relative sensitivity and specificity of drug therapy inappropriateness measures being used by mandated state Medicaid programs to fulfill their OBRA 1990 requirements.

In summary, the results of this project are important because:

- Substantial resources are being spent on outpatient DUR screening without knowing if it validly identifies drug therapy inappropriateness.
- They quantify the relative sensitivity and specificity of DUR screening to detect drug therapy inappropriateness criteria.
- They help to interpret the three-year HCFA-sponsored prospective DUR demonstration project in Iowa.
- With the possibility of a Medicare Drug Benefit in the future, results of this study will be helpful as changes in health care policy further increase the scope of application of outpatient DUR.

Overview of Research Methods

Medicaid patients with evidence of a diagnosis of hypertension were identified from primary medical record data abstraction. Secondary administrative claims data on these patients were available on tape from the state of Maryland's Department of Health and Mental Hygiene. This primary and secondary data sets were used to construct a profile for each patient.

First, a panel of experts in hypertension were asked to agree on explicit criteria for judging hypertensive drug therapy inappropriateness. Agreement was accomplished using a Delphi technique. Second, three physician/pharmacist clinician pairs were recruited and trained to apply the explicit DUR criteria to a study subject profile containing primary clinical data and secondary claims data. In addition to training and the use of forms with which to judge drug therapy inappropriateness and to indicate which criteria for inappropriateness were met, we developed a method for adjudicating disagreement between the clinician pairs. The process of explicit clinician review of the profiles containing primary and secondary data was termed the INDEPTH assessment.

Simultaneously, the explicit DUR screening criteria were applied to the secondary data set of our cohort's Medicaid service and prescription claims to screen for drug therapy inappropriateness. This involved developing a computerized algorithm that applied the intent of the explicit criteria as accurately as possible and was labeled the DURSCREEN assessment.

Third, the results of application of inappropriateness criteria by DURSCREEN were compared to identification of drug therapy inappropriateness found during the INDEPTH review.

Fourth, the relative "specificity and sensitivity" of individual DURSCREEN criteria elements (singly and in combination) were determined by comparing the results of applying them to the clinicians' INDEPTH inappropriateness evaluation. We employed contingency table analysis, construction of receiver operating characteristic (ROC) curves and multivariate regression techniques.

Fifth, blood pressure means of subjects with appropriate drug therapy were compared to blood pressure means of subjects with inappropriate drug therapy as identified by DURSCREEN to test the hypothesis that drug therapy appropriateness is associated with a positive patient outcome.

Lastly, a manual of operations (Appendix A) was developed that can be made available to outpatient DUR programs throughout the country on how to assemble a minimal data set to permit an ongoing assessment of DUR screening of Medicaid claims data when applied to other drugs, diseases and populations.

Statistical analyses were conducted using SPSS™ for Windows version 6.1.2 (SPSS™, 1995) and SAS® version 6.09 (SAS, 1989).

Operational Definitions

Throughout this report we will use the following terms. The reader may want to refer to this section for clarification.

- *Age* - subject age as of the first date of data collection (i.e., index date) for that subject
- *Blood pressure* - We developed eight patient specific continuous blood pressure measures to use in our analyses. These included:
    - Mean diastolic blood pressure - mean of all diastolic blood pressure readings;
    - Mean systolic blood pressure - mean of all systolic blood pressure readings;
    - Mean of the first and second diastolic blood pressures;

- Mean of the first and second systolic blood pressures;
- Change in diastolic blood pressure - the last diastolic blood pressure reading minus the first diastolic reading;
- Change in systolic blood pressure - the last systolic blood pressure reading minus the first systolic blood pressure reading;
- Percent of uncontrolled diastolic blood pressure readings - the number of diastolic blood pressures > 90 mmHg expressed as a percentage of the total number of diastolic blood pressure readings;
- Percent of uncontrolled systolic blood pressure readings - the number of systolic blood pressures > 140 mmHg expressed as a percentage of the total number of systolic blood pressure readings.

- *Compliance ratio* - The compliance ratio was adapted from Farmer (Farmer, Jacobs and Phillips, 1994). Each subject's compliance ratio (COMRATIO) was calculated using the following formula:
  COMRATIO = the mean of the subject's RATIO values. A RATIO was calculated for each antihypertensive drug, for which the subject received at least two prescriptions:

  RATIO = CUMDS/ELAPSED
  CUMDS = sum of days supply of all prescriptions for each antihypertensive drug minus the days supply of the last prescription for the drug.
  ELAPSED = date of the last prescription minus date of the first prescription.

  A compliance ratio of one indicated perfect adherence (i.e., no over- or under-utilization); a compliance ratio less than one indicated under-utilization; a compliance ratio greater than one indicates over-utilization.

- *Criteria* - predetermined elements of health care developed by health professionals relying on professional expertise, prior experience, and the professional literature, with which aspects of the quality, medical necessity, and appropriateness of a health care service may be compared (U.S. Code of Federal Regulations §466.1).

- *Drug use review (DUR)* - an authorized, structured and continuing program that reviews, analyzes, and interprets patterns and instances of drug use against predetermined criteria and standards.

- *DURSCREEN assessment* - screening claims data with computer-applied algorithms representing drug use criteria. We defined DURSCREEN and nine DURSCREEN derivatives which consisted of various combinations of the screening criteria algorithms. DURSCREEN and each of the nine derivatives were used to define drug therapy inappropriateness:
  - DURSCREEN - identified those subjects with at least one flag for any criteria.

- DURSCREEN(2) - identified those subjects with at least one flag for the over-utilization criteria; no other flags were considered in the definition of inappropriateness. These subjects may or may not have received flags for other criteria.
- DURSCREEN(3) - identified those subjects with at least one flag for the under-utilization criteria; no other flags were considered in the definition of inappropriateness. These subjects may or may not have received flags for other criteria.
- DURSCREEN(4) - identified those subjects with at least one flag for any criteria but not including a flag for either under-utilization or over-utilization.
- DURSCREEN(5) - identified those subjects with at least one flag for any criteria but not including a flag for under-utilization.
- DURSCREEN(6) - identified those subjects with at least one flag for any criteria but not including a flag for over-utilization.
- DURSCREEN(7) - identified those subjects with at least one flag for the over-utilization or under-utilization criteria; no other flags were considered in the definition of inappropriateness. These subjects may or may not have received flags for other criteria.
- DURSCREEN(8) - identified those subjects with at least one flag for the dose criteria; no other flags were considered in the definition of inappropriateness. These subjects may or may not have received flags for other criteria.
- DURSCREEN(9) - identified those subjects with at least one flag for the drug-drug interactions criteria; no other flags were considered in the definition of inappropriateness. These subjects may or may not have received flags for other criteria.
- DURSCREEN(10) - identified those subjects with at least one flag for the drug-disease contraindications criteria; no other flags were considered in the definition of inappropriateness. These subjects may or may not have received flags for other criteria.

- *Element* - a categorization of criteria types. Criteria elements include: therapeutic duplication, incorrect dose, drug-drug interaction, drug-disease contraindication, over-utilization and under-utilization.

- *Flag* - any occurrence of an exception to a specific criterion that is interpreted as indicating inappropriate drug therapy. In this study a flag is associated with a specific drug claim for a specific patient.

- *INDEPTH assessment* - application of explicit (structured) criteria to a database including Medicaid claims data *and* abstracted subjects' medical records. In addition, the INDEPTH assessment also includes a "bottom line" implicit assessment of drug therapy inappropriateness. The goal of the INDEPTH assessment is to validate drug therapy inappropriateness identified by the DURSCREEN assessment.

- *Receiver-operating-characteristic (ROC) curves* - an estimate of various combinations of true positive and false positive rates that occur when one uses different methods to, in our case, screen for drug therapy inappropriateness. ROC curves can be used to analyze methods of determining inappropriateness and for estimating the effect on the true positive and false positive rates for different cutoff values (or standards).

- *Sensitivity* - The *sensitivity* of a test is the percentage of individuals with the condition (i.e., drug therapy inappropriateness) who are classified by the test as having the condition.

- *Specificity* - The *specificity* of a test is the percentage of individuals without the condition (i.e., drug therapy inappropriateness) who are classified by the test as not having the condition.

- *Standards* - the degree of acceptable variation from a criterion.

- *Validity* - the degree to which the results of a measurement correspond to the true state of the phenomenon being measured.

    - *Content validity* - the extent to which a particular method of measurement includes all of the dimensions of the construct being measured, and nothing more.

    - *Construct validity* - Given hypotheses about the relationships of a variable, to others being measured in a study, *construct validity* examines whether and how many of the relationships predicted by these hypotheses are empirically borne out when the data are analyzed. For example, if we hypothesize that drug therapy inappropriateness can be more correctly measured with an INDEPTH assessment using medical record *and* administrative claims data *and* implicit *and* explicit criteria (INDEPTH assessment) versus screening of claims data (DURSCREEN assessment), then construct validity would be established if the results from the DURSCREEN assessment vary according to drug therapy inappropriateness as measured by the INDEPTH assessment.

    - *Criterion validity* - the extent to which the measure predicts or agrees with some criterion of the "true" value of the measure.

DISCUSSION

Study Population

Inclusion and exclusion criteria

Our study population was Maryland Medicaid recipients with hypertension. We chose hypertension as a disease to study the validity of a test of medication inappropriateness. Hypertension is a suitable disease to use as an example because:

- Hypertension is a prevalent disease in the general population, and is the most prevalent disease in our cohort of mostly (89%) black people . The reported prevalence of hypertension in Maryland is 20% (26% in black people). In fiscal year 1992, the Maryland Medicaid program had 510,000 recipients; of those, 290,810 (57%) were black.

- Guidelines have been adopted for the diagnosis and treatment of hypertension (1992 Joint National Committee).

Study subject eligibility criteria included: age at least 18 years; a diagnosis of hypertension was noted in the medical record or claims data during the study period; continuous enrollment in Maryland Medicaid without a lapse of eligibility of greater than 31 days during the study period. Subjects were excluded if their medical records were not available for data abstraction or if their prescription claims data were not available.

Sampling frame

Our sample was selected from a cohort of patients from an existing study in which we had encoded and computerized selected primary medical record data variables. This data set was available from another project involving a medication intervention by pharmacists (Metge *et al.*, 1994). It included more than 3,000 Medicaid patients enrolled in four Baltimore adult ambulatory care clinics whose most frequent diagnosis is hypertension. The general medicine clinics were located in inner-city teaching hospitals and the primary physician providers were internal medicine interns and residents, supervised by attending physicians. To remove the effect of this intervention, each subject was assigned an index date that was before the occurrence of any pharmacist intervention. Claims data and medical record data were collected for nine months and six months, respectively, before the index date. The index date for any subject was defined as the first date that medical record data was collected on the subject by a pharmacist. Six months of medical record data were abstracted for each subject backwards from their index date. Nine months of claims data were compiled for each subject prior to the index date. We compiled the additional three months of claims data in order to capture all prescription claims the subject had during the six months of medical record abstraction,

since Maryland Medicaid allows dispensing of 100 days supply of medications used chronically, including many antihypertensive medications. Therefore, although the period of observation was the same for each subject, the index date for subjects varied between April 1, 1993 and January 31, 1995.

Primary Data Collection

Since all data were derived from existing sources, no direct subject contact was necessary. All subject data were confidential and no identifying information could be linked to any study subject. The use of these data was approved by the University of Maryland at Baltimore Institutional Review Board. Two of the four clinic sites provided the bulk of our study population. Clinic One provided 295 study subjects (40%) and Clinic Four provided 263 of the study subjects (36%). The remaining clinic sites contributed about 12% each to the total number of study subjects (Table 1).

Since photocopying our subjects' medical records was not feasible, we abstracted key elements from the primary medical record of the study subjects. Four research assistants were trained to abstract the key data from the medical records of eligible subjects. To assess the accuracy and completeness of the data collection process, we re-reviewed a random sample of 10% of subjects' medical records. The percent agreement for completeness between *Review 1* and *Review 2* was calculated by dividing the lowest number of data elements collected during a review by the total possible data elements identified between *Review 1* and *2*. The percent agreement for accuracy was calculated by dividing the total number of elements that match in both reviews by the number of data elements common to both reviews.

A total of 84 medical records was checked for accuracy and completeness. The percent completeness for all sites ranged between 91% and 96%. The percent accuracy ranged between 95% and 100%. We judged the accuracy and completeness as acceptable and data collection was determined to be complete.

Description of eligible study subjects

We identified 1,003 Medicaid recipients with a diagnosis of hypertension (the diagnosis was identified from the primary or medical record data set) from our sampling frame. Among the 1,003 patients identified, 793 were continuously enrolled in Medicaid without a lapse of eligibility for greater than 31 days during the study period. Five patients were excluded because we were unable to find their medical records. Thus, 788 study subjects were identified in the hospital-based ambulatory care clinics and medical records were abstracted for each of the subjects. Twelve subjects were excluded from the analysis because the diagnosis of hypertension was made after the study period ended or the diagnosis of hypertension was not included in the INDEPTH profiles due to a data entry oversight. Twenty-eight subjects were excluded from the analysis because their

prescription claims data were unavailable. These subjects were Qualified Medicare Beneficiaries and, therefore, also eligible for Medicaid benefits. However, we were unable to find their prescription claims data. Ten were excluded because we failed to delete duplicate prescription claims. Thus, 738 subjects were available for analysis. Of these, 100 subjects could not be classified as having either appropriate or inappropriate antihypertensive drug therapy using the INDEPTH assessment. Therefore analyses comparing the DURSCREEN assessment to the INDEPTH assessment were based on 638 subjects. Descriptive statistics and histograms of our major study variables are presented in Table 2 and Figures 1 through 12.

Most subjects were black (89%) and female (78%). The mean age of this population was 60.6 years (Tables 2 and 3).

The median number of diagnosis categories abstracted from charts and claims data was six. Diagnoses were categorized by the seventeen Classifications of Diseases and Injuries of the ICD-9-CM (International Classification of Diseases, 9th Revision, Clinical Modification, 1993). While several diagnostic categories were more prevalent in females than males, no overall differences were found in the number of diagnoses according to the four sex-race subgroups.

During the six-month period of observation, 99% of the study subjects had at least one prescription claim and 74% had at least one laboratory value. The percent of study subjects with specific types of information from the medical record and the claims data are summarized in Table 4.

The median number of unique antihypertensive drugs per subject was two (Table 2). The range was zero to eleven antihypertensive drugs; 8.4% of subjects had no antihypertensive drugs while 11.1% were prescribed four or more different antihypertensive drugs, although not necessarily concurrently (Table 5). Calcium channel blockers and diuretics were the most frequently used drugs (Table 6).

Twenty subjects did not have a diastolic blood pressure reading and nineteen did not have a systolic blood pressure reading within the six months observation period (one subject was so obese the clinic was unable to record a diastolic blood pressure reading). Eight-eight subjects had a single diastolic and systolic blood pressure reading. The number of blood pressure readings ranged from zero to twelve, the median was three and the mode two.

Four measures were developed for each subject's recorded diastolic and systolic blood pressure readings: mean of all readings, average of the first and second reading, change (the last reading minus the first reading), and the percent of uncontrolled readings. Descriptions of the measures can be found in the operational definitions section and the descriptive statistics in Table 2 and Figures 5 through 12.

Three of the blood pressure measures reduced the number of subject's blood pressure readings over time into a single value. The "percent uncontrolled" measure was an attempt to capture the variations in blood pressure readings recorded within the cross-sectional data. The blood pressure readings were compared to the definition of uncontrolled blood pressure as suggested by the Joint National Committee (1992 Joint National Committee). The percent of uncontrolled blood pressure readings was defined as the percent of diastolic and systolic blood pressure readings > 90 mmHg and 140 mmHg, respectively. The distributions of these measures were multimodal because of the varying number of blood pressure readings per subject (Figures 9 and 10). All of the diastolic blood pressure readings were *controlled* in 46% of the subjects. All of the diastolic blood pressure readings were *uncontrolled* in 11% of the subjects. For systolic blood pressure readings, 22% of the subjects had *all controlled* and 35% had *all uncontrolled* systolic blood pressure readings. Of the 88 subjects with only a single blood pressure reading, 30 had uncontrolled diastolic blood pressure and 53 had uncontrolled systolic blood pressure.

Establishment of Criteria Content Validity

Draft Criteria and Criteria Elements

A draft set of antihypertensive drug use review screening criteria was compiled by the research team using public domain criteria and criteria approved by the state of Maryland's Drug Use Review Board. Drug classes included were diuretics, beta blockers, angiotensin converting enzyme inhibitors, calcium channel blockers, alpha blockers and other antihypertensive drugs. The criteria included DUR screening criteria elements as defined in Section 1927 (g) of the Social Security Act (Table 7). Pregnancy conflicts were included as a drug-disease contraindication.

Identifying a Panel of Experts

Establishment of content validity of antihypertensive drug therapy criteria was accomplished using the Delphi method (Duffield, 1993; Whitman *et al*., 1990). The Delphi panel was composed of those specialists in hypertension who have contributed to the scientific literature (researchers), interpreted the treatment findings (epidemiologists) and developed practice guidelines (clinical practice specialists). To compose a list of specialists, two published lists of experts were first consulted: the list of participants on the Joint National Committee on Detection, Evaluation, and Treatment of High Blood Pressure (1992 Joint National Committee) (the parent committee, and the subcommittees on pharmacologic treatments and clinical evaluation and public health aspects); and the United States Pharmacopeial Convention Panel on Cardiovascular and Renal Drugs.

An electronic MEDLINE search of these committee participants' publications made it possible to judge whether their expertise was indeed in hypertension. Another

MEDLINE search was done to identify antihypertensive therapy review articles (and their authors), written since 1990 and in peer-reviewed journals. A further search for recently published textbooks on hypertension identified authors of book chapters on specific areas in hypertension like therapeutic management. As well, a list of editorial board members for three journals: 1) *American Journal of Hypertension* 2) *Hypertension* and 3) *Clinical and Experimental Hypertension Journal* was obtained and an individual's membership on a board was added to their "expert source" listing. A total of 79 individuals was identified by this process.

Addresses and telephone numbers were obtained from any one of four sources: a 1992 alphabetical and geographic listing of all licensed physicians in the United States; a 1994 listing of all physicians belonging to a certification specialty Board (e.g., internal medicine, family practice); the American Association of Colleges of Pharmacy 1994 roster; and journal articles.

The Delphi invitation mailing consisted of the following: a letter of invitation, a return FAX or mail-in form, summary of the research and a sample Delphi judgment form.

The Delphi invitational mailing was sent initially to 30 of the 79 experts identified. Twenty experts showed a willingness to participate in the Delphi. Fifteen completed and returned the Delphi surveys (Table 8).

Process for Consensus

Most state DUR programs have limited resources to convene an expert panel, (usually from diverse geographic locations) that would meet the requirements for an expert as outlined in the previous section. Ultimately, however, the DUR program's resources are the deciding factor in the choice of a consensus process for the criteria. Having a decision on criteria from a broad perspective is possible by using a Delphi technique. The Delphi technique is a method for overcoming the logistical obstacles presented by a limited budget and geographically scattered experts. The Delphi technique consists of a series of rounds during which a group of individuals is presented with information, usually as statements, and asked to make judgments and supply comments on the items presented. A consensus occurs because the views of the participants converge through a process of informed decision-making.

Originally, the main goal of the Delphi was to improve on the committee process for arriving at a decision by subjecting the views of individual experts to each other's criticism in a way that avoided the psychological drawbacks associated with face-to-face confrontation. Discussion at a committee meeting is replaced by exchanging information through first, a survey administered to the experts by mail asking for their judgment,

followed by surveys containing the outcome of previous survey's (or rounds) judgments. Four rounds or surveys are usually sufficient to arrive at a consensus.

In Round #1, the expert panel was sent 114 screening criteria, covering more than four antihypertensive drug classes, in a booklet form and given the following instructions:

"As you read and evaluate the following drug use review screening criteria, please consider the application of each criterion's statement to the drugs listed in association with it. Screening criteria are computer-applied rules based on readily available outpatient prescription and patient information. They are designed to identify prescriptions that are likely to not conform to the criteria. Screening criteria accept the possibility of false positives and false negatives to increase the efficiency of the criteria application process. The criteria are categorized into at least one of nine types:

(1) adverse drug-drug interaction
(2) drug-allergy interaction
(3) drug-disease contraindication
(4) incorrect drug dosage
(5) incorrect duration of drug treatment
(6) over-utilization
(7) pregnancy conflict
(8) therapeutic duplication
(9) under-utilization

In this document following each criteria statement, you are given three choices with which to evaluate the criteria as well as an opportunity to change or comment on the criteria. The following example explains each of the choices you have to assist you in evaluating the criteria statement. A sample page from the Delphi with simulated comments from a reviewer is attached to these instructions."

Each expert panelist was asked for their judgment about each criterion. Specifically, the panelist marked one of the following choices for each criterion (predetermined element):

(1) Accept criteria as is	Panelist marked this choice if they AGREED that the screening criteria should be applied to all of the drugs listed
(2) Do NOT accept the criteria	Panelist marked this choice if they DISAGREED entirely with the screening criteria.

(3) Unable to evaluate	Panelist marked this choice if they believed that the criterion was outside their domain of expertise and were unable to evaluate it expertly.
(4) Accept criteria with the following modifications	Panelist marked this choice if they disagreed with the inclusion of one of the drugs or its criterion, and listed their reason(s) for disagreement or the drug(s) which they felt should not be included as a part of this criterion.

Before each round was mailed, a number was placed on each Delphi Evaluation Instrument so that returns remained confidential during the analysis process. From the time of mailing, expert panel members were given 30 days to judge the booklet of initial DUR screening criteria. Criteria that either obtained less than 80% acceptance rate in the first round or criteria that had significant modifications suggested by the panel were the only criteria included in Round #2. The acceptance rate was defined as the proportion of those choices that accepted the criteria (with or without modification) to all but the "unable to evaluate" choices:

$$\frac{(1)+(4)}{(1)+(2)+(4)} * 100 = AcceptanceRate$$

The second and final round of the Delphi technique included only criteria that received either less than 80% acceptance rate by the experts or that had significant modifications suggested. To vote, the expert panelists were given only those choices used (marked) to indicate their judgment in round #1. For example, if no respondents marked the choice, "Unable to evaluate," then this choice was not offered for round #2. Several Delphi answer samples were given to the panelists to help them make judgments regarding the criteria presented in round #2.

Results of the Delphi Process

A total of 114 criteria was sent to the expert panel for evaluation and validation. The number of criteria requiring judgment by the expert panelists decreased substantially from Round #1 to Round #2 (114 to 41); 36% of the initial number of criteria required validation in Round #2. Each criterion judged in Round #1 and included in Round #2 included a percentage showing the acceptance rate from Round #1 (with and without modifications). The acceptance rate (expressed as a percentage) is calculated as those experts on round #1 who marked "Accept criteria as is" plus "Accept criteria with the

following modifications" over the total of those who marked any judgment choice other than "Unable to evaluate." The missing response rate averaged 25% across all classes of criteria.

The final consensus of the Delphi rounds is content validation of the DUR screening criteria. Criteria in round #2 of the Delphi technique with an agreement rate of 80% were considered validated for this study (Table 9).

Comparison of the number of criterion per type (e.g., dose, drug-drug interaction, drug-disease contraindication) that were originally proposed against the final criteria revealed that for dose, duplication, duration, over-utilization and underutilization criteria the number remained the same. The greatest reduction in the number of criteria was in the category of the drug-disease contraindication which was reduced from 43 to 27 (63%) specific criteria. The two drug classes affected by the majority of these changes were the beta-blockers and other antihypertensives. Many of the criteria not included in the final criteria were for diseases where the drug was not an absolute contraindication but either required a dose reduction or close monitoring (e.g., beta-blockers used in patients with renal failure, hepatic impairment or in the elderly). The drug-disease criteria were also more likely to be modified and included in Round #2 for re-evaluation. The majority of modifications included the addition or deletion of a drug from that particular drug class or the addition of statements to make the criterion more limited (e.g., presence or absence of a specific disease state or an additional drug). For instance, the drug-disease contraindication of certain beta-blockers in subjects with congestive heart failure was modified to require "not including diastolic dysfunction."

The drug-drug interaction criteria decreased from 40 to 35 specific criteria. Many of these criteria were included in Round #2 requiring modification by the addition or deletion of drugs within that particular drug class. Several criteria were modified to include the absence of a particular disease or concurrent use of other drugs to further restrict the criteria. Finally, of the three appropriateness criteria only one was rejected, the use of certain other antihypertensives as initial monotherapy was "inappropriate."

The Delphi evaluation instrument and content validated criteria applied in this study are included in Appendix B.

DURSCREEN Assessment

The DURSCREEN assessment is an application of DUR screening criteria to claims data using a series of decision rules in a computerized algorithm. Fifty-four distinctive computer-based decision algorithms were used to translate the 92 drug use screening criteria resulting from the Delphi survey for use in the DURSCREEN assessment. Fifty-

three of these algorithms were then used to identify drug therapy inappropriateness for the 738 subjects using Medicaid administrative claims data.[3]

The DURSCREEN system was built upon the existing work already accomplished at the University of Maryland Center on Drugs and Public policy as part of its previous work with the HCFA. Specifically, the Center had previously developed a mechanism for implementing DUR screening criteria using expert systems technology. The decision was made to adapt this system to the needs of the current project.

Assumptions about the Data

Certain assumptions are made about the form and content of the administrative claims data. One assumption is that the drug claim is characterized by the following information:

- Patient ID
- Patient age
- Patient gender
- Physician Identifier
- Provider Identifier
- National Drug Code (NDC) number of the drug
- The date on which the drug was dispensed
- The days supply as reported by the pharmacist
- Quantity dispensed

It was also assumed that the medical service claims data contained the following information:

- Patient ID
- Physician Identifier
- Provider Identifier
- Date of Service
- Diagnostic Code (at least one ICD-9-CM code)

Drug episodes

It is necessary when evaluating drug therapy to determine when a patient begins taking a particular drug, when they stop and if they begin again. Drug claims were

---

[3]One drug-drug interaction criterion (appetite suppressants and guanethidine or guanadrel) was excluded because Maryland Medicaid does not reimburse for appetite suppressants.

organized into drug episodes in order to examine additional characteristics of a course of treatment. Grouping a sequence of prescription claims into an episode of drug therapy captures the natural history of drug use. Two drug claims for a patient were defined as part of the same drug episode if both claims were for the same generic drug and the second claim was dated after the dispensing date of the first and no later than twice the number of days supply after the first dispensing date of the preceding claim.

System Design

The purpose of the screening criteria is to make decisions about whether or not the drug therapy in these claims conforms to the drug use review screening criteria. The process of implementing the drug use review screening criteria as NEXPERT™ knowledge bases consisted of a series of multiple steps of analysis, design, coding and testing. The resulting computer system to apply the criteria was designed as a set of interlocking components that organized the raw claims data and applied the criteria. The flow of information in the system is depicted in Figure 13.

The NEXPERT™ software is an expert systems shell developed by Neuron Data, Inc. for the purpose of formulating and applying sets of related IF-THEN rules (NEXPERT™ 3.0, 1994). In NEXPERT™ terminology, a collection of related rules is a "knowledge base." The software also facilitates the descriptions of entities such as drugs. These descriptions are also considered to be part of a knowledge base. See Appendix A for additional explanations.

Criteria Implementation

The criteria, as established by the expert panel in the Delphi process, were not expressed in a form that could be directly translated into computer algorithms. Additional assumptions and definitions were required so that a form of the criteria as a set of IF-THEN rules could be applied to the administrative claims data. The following describes what additional definitions were required and how the various types of criteria were translated to rules.

Overlapping claims--One of the major decisions in reasoning about drug therapy is to determine when two or more drug claims overlap so that there is an increased probability that the patient is consuming two or more drugs at the same time. In this system two claims are said to overlap when the dispensing date of one is either equal to the dispensing date of the other, or, is later than the dispensing date and earlier than the end date of the other claim. This concept is used to determine likely instances of interaction and multiple doses of a drug prescribed by the physician.

Dose--Evaluation of the total daily dose of the drug prescribed during claim was based on two factors: a) the total daily prescribed dose and b) the maximum allowable

daily dose of the drug as defined in the drug knowledge base. The problem of determining the prescribed total daily dose is somewhat complicated by the fact that a physician may write two or more concurrent prescriptions for the same drug but with different strengths in order to achieve the intended daily dose. For example, if a drug is available in two and three milligram strengths the physician may write two prescriptions for each strength for a prescribed total daily dose of 5 milligrams. We chose to define the total daily dose associated with a given claim as the sum of its estimated daily dose and that of all concurrent claims for the same generic class that were prescribed before or on the same date. This prescribed total daily dose was then compared to the maximum daily dose for a drug as defined by the criteria.

Duplication--The goal of the duplication criterion was to identify instances where a patient was receiving two or more different drugs within the same drug class that was not judged to be therapeutically beneficial. Duplication of the same drug with itself was not included as it was assessed by the over-utilization criterion. The major concern in assessing duplication was to determine when two or more episodes of drug therapy were "overlapping" so that there is an increased probability that the patient is consuming the two drugs at the same time. In particular, we were concerned with the situation where a physician decides to discontinue one course of therapy and begin another before the days supply of the previous claim had expired. To decrease the number of false positives it was decided that for drug therapy to be concurrent both episodes of drug therapy must provide evidence of continuing therapy. Given this decision, for duplication to exist both drug therapy episodes must overlap and consist of a minimum of two prescription claims. The only exception to this rule was when the dispensing dates of both drug claims were on the same day. In this case it was assumed that both drugs were being consumed concurrently regardless of whether or not there was continuing therapy.

Drug-Disease Contraindications--A claim was flagged for a drug-disease contraindication when a medical service claim was found with an ICD-9-CM code that corresponded to the contraindicated disease and the date of service of that claim was prior to or equal to the dispensing date of the claim.

Pregnancy Contraindication--A pregnancy contraindication was considered to be a special case of a drug-disease contraindication. The approach taken was to infer pregnancy as of the dispensing date of a drug claim. The inference was based on the fact that pregnancies do not last for more than ten months and used three signaling events: a) medical services indicating pregnancy, b) medical services indicating pregnancy termination or completion and c) drug claims for prenatal vitamins. A female was considered to be pregnant at the date of dispensing if there was at least one of the three signaling events within the previous ten months and the latest of those was not a pregnancy termination or completion.

Drug-Drug Interactions--A drug claim was flagged as an interacting drug when

a) there was at least one overlapping claim for a drug that interacts with the criteria drug and b) a subsequent claim for the criteria drug existed. The reasoning behind this decision was that a potential drug interaction can occur when a single prescription for an interacting drug is written during the span of time covered by the claim's drug episode. This rule is a deliberate balance between two situations: a) an interacting drug is prescribed without realization of the interaction potential, and b) there is a realization that the drugs interact so that the use of one drug is temporarily suspended or c) there is a short trial to determine if the drugs will lead to an adverse effect for that patient. The decision was made to err on the side of caution and specify that the drug interaction criteria would be violated when an interacting drug was prescribed in a manner that overlapped with a continuing course of therapy regardless of whether or not the interacting drug was continued.

Over-utilization--The goal of the over-utilization or early refills criterion was to identify those cases where the patient was consuming a drug at a higher rate than prescribed. A drug therapy episode indicated overuse of a drug when three conditions were satisfied. First, the drug therapy episode must have a minimum of two claims. Second, the total daily dose of the subsequent claim in the episode must be less than or equal to the total daily dose of the previous claim. The rationale for this rule was: if the physician chose to increase the daily dose, then it was expected that a subsequent prescription for the drug may be "early." Third, the date dispensed of a subsequent claim within an episode must have occurred before the date defined by the previous claim's date dispensed plus 75% of its days supply. That is, over-utilization was said to occur when a new or refill claim for the same drug was dispensed before 75% of the days supply of the previous prescription was exhausted.

Under-utilization--A similar assumption was made about drug episodes for determining underuse. A claim was flagged for underuse when three conditions were satisfied. First, there had to be a previous claim in the same drug episode. Second, the total daily dose of the claim had to be greater than or equal to the total daily dose of the previous claim. The rationale for this was that if the physician chose to decrease the dose of the drug, then it was expected that the refill of a prescription might be late. Third, the claim's dispensing date had to occur after the dispensing date of the previous claim plus the days supply plus ten days. That is, under-utilization was said to occur when the prescription was ten or more days late.

The Criteria Application Process

The criteria application process required that all claims data had been processed to create daily doses for each claim, associate claims into drug episodes, and to order the data within each patient in chronological order.

To begin the evaluation process for a subject all of their drug and medical service claims were retrieved from a file. Next, the first of the chronologically ordered antihypertensive drug claims was selected and identified as the focus claim for evaluation. The system selected the relevant rules from the knowledge base and applied them in the context of the rest of the drug claims and medical services to determine if this claim conformed to or deviated from the criteria. Flags were assigned based upon these evaluations to the focus claim. The system then proceeded to select the next antihypertensive drug claim in chronological order and repeated the process. When there were no more such claims, the evaluation was complete and the resulting claims and flags were written to a file for further analysis.

Rules Development

The NEXPERT™ drug knowledge base for each of the drugs was developed using the criteria developed from the Delphi survey (Appendix B). The rules were subjected to several different testing regimens in order to insure that they correctly applied the criteria. The first step in the testing involved examination of each of the common rules sets to determine if they behaved as intended. Once this was completed and the rules were modified to perform as required, test data sets were developed for every drug in each of the groups of antihypertensive agents. Next, a single unified test data set was developed for a single patient that included drugs from all of the different groups. Finally, the rules were applied to the patient data, and sets of resulting flagged patients were reviewed by the project staff to insure correct application of the criteria. The rules were revised based on several rounds of this last activity until no further errors were detected.

Results

Table 10 presents the DURSCREEN's assessment of each subject's antihypertensive drug therapy as "appropriate" or "inappropriate." Subjects classified as "not appropriate" had one or more flags. Subjects classified as "appropriate" by the DURSCREEN had no flags associated with their drug therapy. Table 11 lists the total numbers of subjects by six criteria elements classified as "inappropriate." The dose, duplication, over and underutilization criteria elements were each defined by a single algorithm. The drug-diseases contraindication and drug-drug interaction criteria elements included twenty-one and twenty-eight different algorithms, respectively.

The DURSCREEN assessment identified 63.1% of the 738 subjects as "inappropriate." In comparing Table 10 and Table 11 it appears that most of the "inappropriate" subjects were identified by the under-utilization criterion (345 out of 466 or 75% of the "inappropriate" subjects).

Tables 12 through 15 provide additional comparisons of the DURSCREEN results. Table 12 is an expansion of Table 11 and presents the number of subjects that failed each

of the specific criteria and the flag frequency for that criterion for each subject. A flag denotes that a subject has failed a criterion. A subject may have received more than one flag per criterion if the subject received more than one antihypertensive drug. A flag is assigned to a subject and an antihypertensive drug. For example, if a subject is receiving three different antihypertensive drugs each drug may potentially fail the dose criterion for a maximum of three flags.

Twenty-three of the 53 criteria rules (43%) generated a flag. As expected, most of the criteria that did not generate a flag were for rarely prescribed drugs such as guanethidine and guanadrel. Fourteen of the 15 subjects that failed the drug-drug interaction between potassium-sparing diuretic and angiotensin converting enzyme inhibitors received two flags: one flag for the potassium-sparing diuretic and one flag for the angiotensin converting enzyme inhibitor. The largest percent of subjects failed the over-utilization, under-utilization and drug-drug interaction criteria elements (23%, 47% and 15% of subjects, respectively).

Tables 13, 14 and 15 classify the subjects by the DURSCREEN assessment of "appropriate" (no flags) versus "inappropriate" (≥1 flags). Table 13 compares the flag frequency per subject between the assessment classifications and includes all 23 criteria for which flags were generated. Table 14 excludes the flags generated by the over-utilization and under-utilization criteria. A comparison of Tables 13 and 14 shows that 284 (61%) subjects classified by DURSCREEN as "inappropriate" were so classified because they failed the under-utilization or over-utilization criteria only. Table 15 compares the number of unique criteria each subject failed by the DURSCREEN assessment. A total of 201 subjects (43%) classified by DURSCREEN as "inappropriate" failed greater than one criterion.

In summary, based on a series of 53 decision rules, drug therapy inappropriateness for the 738 study subjects was identified using data from the Medicaid claims database. A single instance of any flag for a subject was considered to indicate inappropriate therapy. Nearly two-thirds of all study subjects were identified as receiving inappropriate drug therapy. Utilization (both over-utilization and under-utilization) was the primary identifier for identifying drug therapy inappropriateness. DURSCREEN derivatives (that is, different combinations of the computer-based rules) were developed and explored. The number of DURSCREEN flags ranged from zero to ten, the mode was zero and the median was one flag per subject. The median and mode number of criteria elements was one per subject.

INDEPTH Assessment

Description of the INDEPTH Assessment

The INDEPTH assessment of antihypertensive drug therapy inappropriateness was developed to approximate a "gold standard" measure of inappropriateness. To facilitate this "closer to the clinical decision" INDEPTH assessment, Medicaid hypertensive patient profiles were built from several information sources. Primary data from four hospital-based ambulatory clinic sites were abstracted from medical records and secondary data from the administrative claims database were combined to "build" each subject's profile. We describe the process of INDEPTH assessment measure development and provide results that we used to validate the INDEPTH assessment as a measure of drug therapy inappropriateness for subjects with hypertension.

Profiles

All of the information about each subject was computerized and formatted in a standard medical profile format. Thus, the reader was unable to identify the source of any subject information because of the standard medical profile format used. Some subject profiles were lengthy and consisted of several pages; others were brief and consisted of the minimum of two pages. A simple font was used to print all data entries. All data entries were grouped into five categories: diagnoses, medications, laboratory data, physical findings and procedures. These data were printed and presented in chronologic order. As described below, a subject profile and assessment instruments were distributed for review by our expert panel of physicians and pharmacists. These data provided the basis for the INDEPTH assessment of drug therapy inappropriateness. A sample subject profile is included in Appendix C.

We developed data assessment instruments that were specific to each drug and a global form to record antihypertensive drug therapy inappropriateness. The drug-specific forms included the explicit criteria developed from the Delphi survey. The global assessment form prompted the reviewer to assess the subject's antihypertensive drug therapy as "appropriate," "inappropriate," or "cannot determine." Copies of these forms can be found in Appendix D.

Reviewers

Our expert panel of reviewers consisted of three physicians and three clinical pharmacists. Selection was based on either their expertise or extensive clinical experience in the management of hypertension or the principles of DUR. We trained the reviewers on three separate occasions interspersing several practice sessions using ten subjects each. We presented a "mock" adjudication panel that simulated the process of achieving a consensus. Appendix A contains details about selecting and training the

expert panelists. The flow of information for the INDEPTH assessment is presented in Figure 14.

Review Process

Each subject was randomly assigned to two reviewers, a pharmacist and a physician. The subject's profile was independently reviewed and the results were recorded on the appropriate assessment instruments. Batches of forty or more profiles were distributed at distinct time points (seven overall) and reviewers were required to return their assignment within a specific time. Three review types were specified as follows: initial review (which provided the basis for our INDEPTH assessment), inter-rater review (which provided a companion measure from a distinct physician-pharmacist pair) and an intra-rater review (which provided a reassessment by the same physician-pharmacist pair at a point later in time). Each panelist was unaware of the review type of any given profile evaluation but all panelists were aware that we were conducting some quality control measures using re-reviews. Each panelist assessed between 343 and 345 profiles.

Reviewers were required to assess each subject's profile and record their overall drug therapy appropriateness assessment as follows: appropriate, inappropriate or cannot determine. An "initial" assessment was undertaken by a physician-pharmacist pair. The INDEPTH assessment of inappropriateness was established when the physician-pharmacist pair agreed (that is, the "initial" assessment became the "final" assessment). When the two assessments did not agree, the subject's profile was referred to a consensus panel for a second, final assessment; i.e., adjudication and assignment of an INDEPTH assessment. The consensus panel consisted of at least four of the six reviewers. At least two pharmacists and two physicians were required to be present.

A consensus panel was convened as the roster of unadjudicated profiles grew. When the panel reviewed each profile collectively, they had an opportunity to discuss and debate the relative merits of the subject profile including insufficient data. We provided decision rules when the consensus panel could not agree on drug therapy appropriateness. Appendix A contains a detailed description of the review and consensus process.

To assess various aspects of quality control, a 25% sample of subjects was randomly assigned for intra- or inter-rater reliability. Our rater reliability results were similar to those reported in the literature (Coulter, Adams and Shekelle, 1995; Localio *et al.*, 1996). The percent agreements ranged from a low of 65.1% to a high of 81.0%. Intra-rater agreement was much better than inter-rater agreement. The overall intra-rater agreement for the physicians was substantial at 81.0% [Kappa 0.57 + (0.09)] and for the pharmacists, moderate 73.4% [Kappa 0.49 + (0.09)]. Inter-rater agreement was 74.4 % [Kappa 0.39 + 0.08] for physicians and 65.4% [Kappa 0.32 + 0.08] for pharmacists. Additional details of these findings are reported in Appendix E.

A profile identifier (ID) was assigned to each subject unique to the panelist and the review type. Therefore, each subject was assigned a minimum of two profile IDs for a primary review and four profile IDs if the subject was also reviewed for inter-rater or intra-rater reliability. A total of 2,062 profile IDs was generated (#0001 to #2062). The profiles and assessment forms were collated and rechecked before distribution to the panelists.

Results

INDEPTH assessment was completed for 788 subjects; 738 were eligible for final analysis. The initial physician and pharmacist readings by the INDEPTH assessment are shown in Table 16. Seven hundred-thirty eight subjects are shown according to three categories: appropriate drug therapy, inappropriate drug therapy and cannot determine. An overall initial agreement rate of 65.1% is shown by summing up the entries in the last column of the data table. One-third of all subjects were sent to the panel for adjudication. One hundred of the 738 subjects were labeled as "indeterminate" when they could not be classified as having either appropriate or inappropriate antihypertensive drug therapy using the INDEPTH assessment. Of the remaining 638 study subjects, 155 subjects (24.3%) were identified as having inappropriate drug therapy.

Table 17 compares the diagnostic groups among the appropriateness determinations by our expert panel. Panel determinations were not influenced by the number and type of diagnostic groups. The proportion of subjects are shown in each of the diagnostic groups. Each subject had one or more diagnosis and all subjects (100%) had a circulatory diagnosis (hypertension). More than 50% had a second circulatory diagnosis besides hypertension. In all but one diagnostic category (mental), no statistically significant differences were found. Among subjects with at least one mental diagnosis, a higher proportion of subjects were observed in the "appropriate" category. Since the number of comparisons was large, no significance was inferred from this finding.

Profile of "Cannot Determine" Subjects

The demographic profiles of subjects with appropriate and inappropriate drug therapy were compared with the group of "indeterminate" study subjects. No differences were found for sex, race, age, and the number of disease categories. Seventeen percent of those designated "indeterminate" did not have a single blood pressure reading. Of the remaining 83 subjects, more than 90% had an uncontrolled blood pressure reading but limited blood pressure readings were available to follow the course of the subject during the period of observation. The distinguishing feature for panelists labeling these subjects as "cannot determine" was the lack of laboratory and physical findings data.

Validation of the INDEPTH Assessment

The main validation feature for the INDEPTH assessment focused on blood pressure control. The eight continuous blood pressure measures (mean, mean of first and second blood pressures, change in blood pressure and percent of uncontrolled blood pressure readings) were compared to the INDEPTH assessment (Tables 18 - 21). The group of study subjects with appropriate drug therapy consistently demonstrated statistically significant lower blood pressure measures and demonstrated statistically significant reductions in blood pressure. The percent of uncontrolled blood pressure readings was shown to be statistically significantly higher among the group of subjects identified with inappropriate drug therapy. These findings provide evidence for the validity of the INDEPTH assessment as a measure of antihypertensive drug therapy inappropriateness.

Comparison of DURSCREEN assessment and INDEPTH assessment

The comparison of the basic screening instrument, DURSCREEN, with the INDEPTH assessment findings are shown in Table 22 and demonstrated statistically significant associations but poor agreement (47.9%). The measure of sensitivity was 0.735 compared with a lower specificity (0.395). Alternative DURSCREEN derivatives detailed below demonstrate varying levels of agreement, sensitivity, specificity and statistical association. These findings will be presented to show modifications in the ways DURSCREEN could be operationally defined.

Because of the very large proportion of drug therapy inappropriateness identified by DURSCREEN, we operationally defined nine DURSCREEN derivatives using empirical data. These derivatives consisted of several combinations of the computer-based screening algorithm. Table 23 provides a summary evaluation of sensitivity and specificity measures for the various DURSCREEN derivatives compared with the panel determinations. DURSCREEN and the nine DURSCREEN derivatives [DURSCREEN(2) - DURSCREEN(10)] are defined in the Operational Definitions, and will be reviewed here. DURSCREEN(2) identified those subjects with at least one flag for the over-utilization criteria; no other flags were considered in the definition of inappropriateness. These subjects may or may not have received flags for other criteria. Similarly, DURSCREEN(3) included all subjects with at least one flag for the under-utilization criteria. DURSCREEN(4) operationally defined inappropriateness as at least one flag for any criteria without consideration for the over-utilization and under-utilization criteria. DURSCREEN(5) consisted of all DURSCREEN criteria except under-utilization. DURSCREEN(6) is operationalized with all DURSCREEN criteria minus over-utilization. DURSCREEN(7) combines all of the flags specific to utilization (both under and over utilization). DURSCREEN(8), DURSCREEN(9) and DURSCREEN(10) include subjects with at least one flag specific to dose, drug-drug interactions or drug-disease contraindications, respectively.

In contrast to the original DURSCREEN, DURSCREEN(8) demonstrated the highest specificity (0.903) and overall agreement (73.5%). DURSCREEN(8) was optimal at screening out the subjects with appropriate drug therapy but with a sensitivity of 0.213 performed poorly in identifying drug therapy inappropriateness.

DURSCREEN (9) demonstrated the third highest percent agreement (72.7%) and second highest specificity (0.896). Unfortunately, it demonstrated the lowest sensitivity--a feature that we are clearly trying to optimize. One other derivative, DURSCREEN(5) offered a middle of the ground approach with a 61.9% agreement rate, and measures of sensitivity and specificity of 0.561 and 0.638, respectively.

Refinements in operationalization of the utilization criteria may demonstrate that all DURSCREEN flags and flag elements may be useful, but it was apparent that the "all or none" screening approach, especially with utilization flags, offered little utility.

Receiver-Operating-Characteristic Curves

Construction of receiver-operating-characteristic (ROC) curves is an example of one of the methods available to analyze a situation missing a "gold standard." ROC curves give us a graphical representation of the compromises that could be made between "true" positives and "false" positives. ROC curves derive their name from the description of the inherent detection *characteristics* that the combination of INDEPTH and DURSCREEN (2x2 tables) gives the *receiver* of the DURSCREEN results (e.g., a state DUR Board or a Pharmacy and Therapeutic Committee) to base their decisions (*operate*) at any point on the curve by using an appropriate decision threshold.

ROC curves describe inappropriateness detectability that is independent of prevalence of drug therapy inappropriateness and would be useful for two reasons. The first concerns a decision the DUR Board must make about which criteria to operationalize prospectively (vs. retrospectively)--if one finds a high false positive rate when comparing DURSCREEN with INDEPTH, then prospective review is likely not in order. The second concerns the rate of false positives that are acceptable to the DUR Board -- if the therapy is screened as inappropriate and the effect is life-threatening when it is a true positive, a high false positive rate may be more tolerable in exchange for keeping true positives high and the false negatives low. The comparison of two methods (INDEPTH, DURSCREEN) of detecting drug therapy inappropriateness, also called convergent construct validity, would give us a better understanding of how DUR screening works in relation to having more comprehensive clinical information on a Medicaid hypertensive subject.

ROC analysis was used in an attempt to "improve" the statistical relationship between DURSCREEN and the INDEPTH findings. A ROC curve is constructed by plotting the true positives (y-axis) by the false positives (x axis) for various conditions, or cutoff

points. The cutoff point with the highest true positive value and lowest false positive value (that is, the point in the uppermost left corner of the graph) is the condition with the highest sensitivity and specificity. We constructed the ROC curves using the exponential model described by England (England, 1988).

The number of DURSCREEN flags and the number of different types of flags was explored. Figures 15 through 18 show ROC curves generated for four scenarios. Sensitivities and specificities of the four curves and cutoff points are included in Tables 24 through 27. The characteristics varied in the curves included: total number of flags detected in the DURSCREEN (Figure 15, Table 24); number of criteria elements flags (drug-drug interaction, drug-disease interaction, therapeutic duplication, dose, over-utilization, under-utilization) detected in the DURSCREEN (Figure 16, Table 25); and total number of antihypertensive drugs (Figure 17, Table 26); total number of flags (excluding utilization flags) detected in the DURSCREEN (Figure 18, Table 27).

Although all ROC curve areas-under-the-curve (AUCs) were statistically different from chance occurrence (i.e., AUC equal to 0.5), the magnitude of difference from 0.5 was small (AUCs ranged from 0.6011 to 0.6568) and offered little utility. The height and skewness of the curves provided little assistance in selecting a cutoff for increasing sensitivity and specificity of the DURSCREEN. The most improvement in sensitivity and specificity observed was for a cutoff of two or more antihypertensive drugs (sensitivity 0.794, specificity 0.431). Unfortunately, this model does not incorporate the criteria elements used in the DURSCREEN assessment.

Relationship Between DURSCREEN and Blood Pressure

To test our hypothesis (i.e., subjects with appropriate antihypertensive drug therapy have lower mean blood pressures than subjects with inappropriate antihypertensive drug therapy), we estimated our minimal sample size to be ten per group based on the following parameters. We anticipated that appropriately treated subjects will have lower mean blood pressures. We desired to detect clinically meaningful differences (8 mmHg for diastolic blood pressure, 12 mmHg for systolic blood pressure). We established alpha at 0.05, beta was 0.2 and power was 0.8. Student's t-test was used to test for differences in mean blood pressures between groups.

The mean systolic blood pressure for subjects having "appropriate" antihypertensive drug therapy (as determined by DURSCREEN) was 141.9 mmHg (S.E. 1.2) compared to 144.1 mmHg (S.E. 1.0) for the "inappropriate" group. The average diastolic blood pressure for the subjects with "appropriate" antihypertensive drug therapy (as determined by DURSCREEN) was 82.9 mmHg (S.E. 0.7), compared to 82.0 mmHg (S.E. 0.5) for the subjects with "inappropriate" antihypertensive drug therapy. We found that the DURSCREEN did not differentiate blood pressures among the subjects with "appropriate" versus "inappropriate" antihypertensive drug therapy.

Because other variables may influence blood pressure, we developed a series of multivariate models using four continuous measures of blood pressure. Two measures were specific for systolic blood pressure and two were specific for diastolic blood pressure. The operational definitions for each measure have been previously defined (see Operational Definitions) and a brief synopsis is restated with each model description. The development of each model included a single measure (YES/NO) of drug therapy inappropriateness as determined by the computer-based DURSCREEN (the original and nine derivatives) and four control variables identified as clinically and statistically important in model development. These models constitute further testing the hypothesis that subjects with appropriate antihypertensive drug therapy (as identified by DURSCREEN) have lower mean blood pressures than subjects with inappropriate antihypertensive drug therapy.

Control Variables

In the development of the multivariate models, several variables may have direct or indirect influences on one or more of the dependent variables under study. The four control variables are included in all models presented. These are: age, compliance ratio, the number of antihypertensive drugs prescribed and the number of disease categories abstracted from the subjects' medical records and claims data. It would be difficult to learn whether these variables were part of a "causal" chain and our data simply could not demonstrate cause and effect. Evidence that these four variables can influence our dependent variables was supported by the literature (Caldwell *et al.*, 1983; Hawkins, Bussey and Prisant, 1997).

One of our control variables, compliance ratio, was derived from other variables in the Medicaid claims data. The methodology for calculation for the compliance ratio was adapted from Farmer (Farmer, Jacobs and Phillips, 1994). To calculate the compliance ratio the subject had to have received a minimum of two prescriptions for the same drug. Consequently, the compliance ratio could not be calculated for 87 of the 638 subjects used in this analysis because they had less than two claims for any single antihypertensive drug. In the original proposal a subject was required to have a minimum of four prescriptions, however, this rule would have eliminated over 25% of the subjects from analysis. A ratio was calculated for each antihypertensive drug the subject received during the study period. This ratio was determined by summing the days supply for each antihypertensive drug minus the days supply of the last prescription for that drug. This value was then divided by the elapsed time between the dispensing of the first prescription and the last prescription for that drug. The compliance ratio was then calculated by taking the mean of all the ratios for each antihypertensive drug the patient received. The days supply for the last claim was excluded because the elapsed time to the next prescription could not be determined. The mean compliance ratio was 0.84 (S.D., 0.30). Compliance ratios ranged from a low of 0.13 to a high of 4.17. Ten percent of the study population had a compliance ratio of less than 0.50 and 4.4% had a compliance

ratio greater than 1.20. Approximately, 58% of the study population had a compliance ratio greater than .75 and less than 1.10. There were no sex-race differences.

Attributes of the models are presented in four data tables, one for each operational expression of the dependent variable (Tables 28 - 31). For each dependent variable, the first model included DURSCREEN as originally defined and the nine subsequent models included the various DURSCREEN derivatives described earlier. Each model included one of the DURSCREEN measures coded in binary (appropriate=0, inappropriate=1) format and the four additional control variables (all continuous measures), namely, age, compliance ratio, the number of antihypertensive drugs and the number of medical diagnoses. An eleventh model is also presented with the same cadre of independent control variables and the results of the INDEPTH assessment are used in place of the DURSCREEN measure. The INDEPTH assessment of inappropriateness is presented in the last row of each table for comparison with the DURSCREEN measures. Of particular importance is the overall amount of explanatory variance (the adjusted R^2), the significance of each model (F and corresponding level of significance) and the level of significance for each of the predictor and control variables.

Mean of First and Second Systolic Blood Pressures

Ten multivariate models are presented in which the dependent variable was operationally expressed as the mean of the first two systolic blood pressure readings abstracted from the subject's medical record (Table 28). All of the multivariate models were statistically predictive of the mean of the first and second systolic blood pressures with age and the number of antihypertensive drugs prescribed contributing significantly to each model. No single DURSCREEN model emerged as the best model with average explanatory variance about 9%. None of the DURSCREEN measures were statistically related to the mean of the first and second systolic blood pressure readings. The compliance ratio failed to provide any predictive value when all other variables were controlled. The INDEPTH assessment of inappropriateness, with the control variables, accounted for 25% of the explained variation in the model and four of the five predictors were significant ($p<0.05$). The compliance ratio failed to achieve statistical significance but is believed to be important as a control variable.

Mean Systolic Blood Pressure

Ten multivariate models are presented in which the dependent variable was operationally expressed as the mean of all systolic blood pressure readings abstracted from the medical record (Table 29).

All of the multivariate models were statistically predictive of average systolic blood pressure with age and the number of antihypertensive drugs prescribed contributing significantly to each model. No single DURSCREEN model emerged as the best model

with an explanatory variance of 10%. None of the DURSCREEN measures were statistically related to average systolic blood pressure. The compliance ratio and the number of diseases failed to provide any predictive value when all other variables were controlled. The INDEPTH assessment of inappropriateness, with the control variables, accounted for 33% of the explained variation in the model and four of the five predictors were significant ($p<0.05$). The compliance ratio failed to achieve statistical significance but is believed to be important as a control variable.

Mean of First and Second Diastolic Blood Pressures

Ten multivariate models are presented in which the dependent variable was operationally expressed as the average of the first two diastolic blood pressure readings abstracted from the medical record (Table 30). All of the multivariate models were statistically predictive of the mean of the first and second diastolic blood pressures with age being the only predictor contributing significantly to each model. None of the remaining predictors, including all of the DURSCREEN measures, contributed any statistical explanation to the models. The explanatory variance for each model was 3% to 4%.

The INDEPTH assessment of inappropriateness, with the control variables, accounted for 15% of the explained variation in the model and two of the five predictors were significant ($p<0.05$). The compliance ratio, the number of antihypertensive drugs prescribed and the number of diagnostic categories failed to achieve statistical significance for the initial diastolic blood pressure model.

Mean Diastolic Blood Pressure

Ten multivariate models are presented in which the dependent variable was operationally expressed as the mean of all diastolic blood pressure readings abstracted from the medical record (Table 31).

All of the multivariate models were statistically predictive of average diastolic blood pressure with age, compliance ratio and the number of antihypertensive drugs prescribed contributing significantly to each model. No single DURSCREEN model emerged as the best model with an explanatory variance of 5%. None of the DURSCREEN measures were statistically related to average diastolic blood pressure. The number of diseases failed to provide any predictive value when all other variables were controlled. The INDEPTH assessment of inappropriateness, with the control variables, accounted for 20% of the explained variation in the model and two of the five predictors were significant ($p<0.05$). The compliance ratio, number of antihypertensive drugs and number of diagnostic categories failed to achieve statistical significance.

Compendia of Blood Pressure Measures with DURSCREEN Criteria

To determine if any of the criteria were predictive of blood pressure, we developed 14 multiple regression models. Seven models used the mean diastolic blood pressure as the dependent variable and seven models used the mean systolic blood pressure as the dependent variable. All 738 subjects were eligible to be included in these models. Four control variables were included in each model (number of diagnostic categories, the compliance ratio, number of different antihypertensive drug and age). One of seven criteria was included as an independent variable in each model. These seven criteria were chosen because (a) a failure of these criteria may result in a change in blood pressure; (b) there were sufficient subjects eligible for application of the criterion to include in the model[4], and (c) at least one flag occurred for the criterion. The seven drug use criteria selected were:

- dose
- duplication
- under-utilization
- over-utilization
- diuretics and indomethacin drug-drug interaction
- potassium-wasting diuretics and cholestyramine or colestipol drug-drug interaction
- centrally acting antihypertensives (clonidine, methyldopa, guanabenz or guanfacine) and tricyclic antidepressants drug-drug interaction

Thus, 14 models were developed from all possible combinations of the two dependent variables and seven drug use criteria variables. Each of the 14 models included the following: one of the two blood pressure measures as the dependent variable (seven models included the mean diastolic blood pressure and seven models included the mean systolic blood pressure); four control variables (age, number of diagnostic categories, compliance ratio and number of antihypertensive drugs) as independent variables; and one of seven DUR screening criteria as a fifth independent variable. Each criterion was defined in an "all or none" phenomenon (appropriate=0, inappropriate=1) as potential

---

[4]We performed each regression model on only those subjects who were eligible for a criterion. In other words, in order for subjects to be eligible for the dose or duplication criteria, they had to have at least one claim for an antihypertensive drug during the study period. Subjects eligible for the over or underutilization criteria had to have at least two claims for the same drug entity. Subjects eligible for a drug-drug interaction had to have received at least one claim for the antihypertensive drug of interest. Additionally, we excluded subjects from the model if there were insufficient claims to calculate a compliance ratio or if they were missing dependent variable measure (i.e., they did not have a systolic or diastolic blood pressure reading). Consequently, the number of subjects eligible for each model varied from 71 to 614.

predictors for the two expressions of blood pressure measurement. All variables were entered in the regression equation.

Regression Models

The variables were examined for outliers (greater than three standard deviations) and influencing data points were evaluated using Cook's distance (SPSS™, 1995). Review of the correlation matrix for all 14 regression models revealed that none of the control or criteria variables were highly correlated with either the mean diastolic or mean systolic blood pressure; correlation coefficients ranged from -0.21 to 0.37.

Although several of the correlation coefficients between the independent variables were statistically significantly ($p< 0.05$), the majority of the coefficients were low (-0.12 to 0.31). One notable exception was the correlation between age and the number of diagnostic categories which was moderately correlated (-0.36 to -0.64) across all 14 models.

The beta coefficients and model statistics are reported in Tables 32-38. Thirteen of the fourteen models were significant ($p<0.05$). The exception was the model of mean diastolic blood pressure which included criterion #51 (tricyclic antidepressant and adrenergic agents drug-drug interaction) as an independent predictor variable. Nonetheless, the amount of explained variance for all models was small, with adjusted R^2 ranging from 0.08 to 0.18 for mean systolic blood pressure and from 0.05 to 0.07 for mean diastolic blood pressure. Duplication (criterion #2) and the indomethacin and diuretics drug-drug interaction (criterion #36) were the only criteria variables which achieved significance ($p<0.05$) and only in the mean systolic blood pressure model. Both beta coefficients were inversely related to the mean systolic blood pressure.

Overall, the individual criteria did not provide statistical insight into the expressions of blood pressure assessment. The small amount of variance predicted by each of the models suggests that additional variables are necessary to further explain the dependent variables. For example, many important variables (body mass index, diet, marital status) were not available for this study. However, the limited variance explained by the these criteria models suggests that "inappropriate" therapy as defined by lack of conformance to a single criterion is minimally related to blood pressure measurement.

Limitations

We have identified several limitations to our findings. First, our study focused only on inappropriateness of antihypertensive drug therapy. Generalizing our findings to inappropriateness measures for drug therapy of other diseases would be premature. However, our study offers the framework for reproducing the methodology to other

diseases, such as asthma, hyperlipidemia, congestive heart failure or coronary artery disease.

Our cohort does not represent the general population. There is an over-representation of black females and the cohort is from the Medicaid population and the cohort is drawn from the Maryland Medicaid population treated in hospital-based clinics. Another important limitation is that this was a cross-sectional study, and was not designed to measure outcomes of inappropriate drug therapy. Because of the cross-sectional observational design, not all subjects were evaluated using the same amount of data. Although the period of observation was the same for each subject (six months), the amount of data such as blood pressures and laboratory values was dependent on other factors. Specifically, the amount of data was determined by the number of primary care visits the subject encountered during the study period. We did not attempt to "weight" the value of subject data based on the number of clinic visits.

The cross-sectional design of our study limited our ability to collect additional variables that may have improved the predictability of our regression models. For example, many important variables (body mass index, diet, marital status) were not available for this study. Additionally, we used mean blood pressure measurements as our dependent variable, which did not allow us to examine the temporal relationship between changes in blood pressure and the presence of inappropriate drug therapy. Given these limitations, however, the models strongly suggested that individual drug use criteria are poorly related to blood pressure.

Not all study subjects could be evaluated by our expert panel. Missing data or scant data regarding blood pressure measurements prevented 100 study subjects from being evaluated and their drug therapy appropriateness was subsequently designated as "cannot determine." Interestingly, the computerized algorithms (DURSCREEN) identified the same proportion of appropriate and inappropriate among the "cannot determine" category as found among subjects where appropriateness could be determined using the INDEPTH assessment.

We developed and employed a measure of drug therapy inappropriateness, previously untested but taking into account a myriad of limitations. We incorporated clinical measures (e.g., blood pressures) into our evaluation instruments used to assess appropriateness. It is possible that the physician and pharmacist reviewers were basing their appropriateness judgment on the characteristics of the blood pressure values and not necessarily on any characteristics of drug therapy. However, to diminish this possibility, we trained the reviewers to use the explicit drug use criteria developed in our Delphi survey, and required them to use assessment instruments that forced them to consider the content validated drug use criteria in their assessments.

We did not rely on a single individual in making the determination of appropriateness. We required agreement about appropriateness and inappropriateness between a physician and a clinical pharmacist. Where disagreements occurred, we employed specific adjudication approaches to resolve differences. Further, we acknowledged that not all subjects could be adequately assessed.

Additionally, blood pressure measurement readings were not taken in a controlled environment and, therefore may be inconsistent. We had little knowledge about equipment calibration and limited knowledge about how the blood pressures were taken (e.g., standing, sitting). It is these blood pressure readings, however, on which antihypertensive prescribing decisions were made for our study subjects. We therefore did not attempt to determine the validity of the actual blood pressure readings.

Our assumption that the INDEPTH assessment is the closest we have to a gold standard is a limitation, especially in the interpretation of the results of our study. One could argue that, although our INDEPTH assessment was statistically and clinically associated with effectiveness (blood pressure), we did not attempt to evaluate any association with adverse drug therapy outcomes. To demonstrate this relationship, a prospective, longitudinal study design should be employed, since adverse events are relatively rare. A prospective design would allow collection of necessary data (e.g., serum drug concentrations) to identify whether a drug-drug interaction resulted in an adverse event. A longitudinal study would give one the opportunity for a longer observational period to capture true incidence rates of clinically significant adverse drug therapy events. Despite these limitations, the INDEPTH assessment has utility as a measure of "truth" in the assessment of drug therapy inappropriateness.

CONCLUSIONS

Summary of Major Results

We developed a drug therapy evaluation tool (INDEPTH assessment) used to assess medication inappropriateness in a cohort of Medicaid hypertensive patients. These tools were used by an expert panel of physicians and pharmacists to determine the inappropriateness of patients' antihypertensive drug therapy. Clinical data taken from patients' clinical-based medical records were abstracted and summarized in an electronic data base. This clinical data was merged with administrative claims data, including medical service and prescription claims. Antihypertensive drug history, laboratory values, diagnostic groupings, blood pressure readings, weight and follow up history were provided to our panel for their review. A final determination was made about each patient's drug therapy appropriateness. These determinations were tested against blood pressure readings to validate our expert panel findings. One out of five of our subjects was identified as receiving inappropriate drug therapy. In a number of subjects (14%), a determination of drug therapy inappropriateness could not be made by our expert panel. We believe that this may be due to insufficient data available, including both the claims data and abstracted medical records.

We also developed computerized algorithms (DURSCREEN assessment) and a minimal data set to assess drug therapy inappropriateness. The computer algorithms used in our DURSCREEN were as specific as technologically feasible, taking into account both temporal relationships and concurrency of events. In many DUR programs, a drug-drug interaction might be defined as having occurred if a patient took both of the involved drugs anytime during a review period (e.g., six months). This is a very sensitive definition -- it will find all the cases of the interaction -- but it is not specific because it will include cases where the drugs were taken many months apart and perhaps never taken concurrently.

The DURSCREEN assessment was applied to study data to decide whether a patient's antihypertensive drug therapy was appropriate according to a series of decision rules. Three out of five subjects were identified as having inappropriate drug therapy using this method.

The INDEPTH assessment findings were compared with the findings obtained from the computerized drug use review algorithms (DURSCREEN). The computerized algorithms failed to provide clear insight into the findings obtained from the richer, clinic database and assessment of inappropriateness by the INDEPTH assessment.

We conclude that the computerized algorithms used to monitor and evaluate the Medicaid database are not sensitive or specific enough to detect true cases of drug therapy inappropriateness. In other words, claims data may not be rich enough to

replicate clinical insight into the patient's medical history for the purpose of establishing drug therapy inappropriateness. It appears that this clinical insight is a prerequisite for assessing drug prescribing offered through the routine claims processing transactions.

Policy Implications

We acknowledge that outpatient DUR programs as mandated by OBRA 1990 legislation were intended to incorporate a *screening* process for *potentially* inappropriate drug therapy. DUR screening is an administrative procedure designed to separate prescriptions into groups of those more likely to present a problem from those less likely to be problematic. Thus, screening involves applying some test to many cases to detect a small number of problems. A good screening test can detect as many as possible of the cases that exist; such a test is said to have high sensitivity.

However, since a positive screen leads to additional work -- diagnostic confirmation, preventive or therapeutic intervention, and follow-up -- a good screening test must also avoid labeling those cases that have no problems as problematic. Such cases are called false positives and tests that have few of them are said to have high specificity. When screening for individual cases of inappropriate therapy, false positives may incur a cost not only in economic terms, but in emotional terms, such as patient worry.

False negative screens (i.e., not detecting a problem when there is one) can also be problematic in DUR screening. For example, some drug-disease contraindications are not reliable in claims data. Our cohort of hypertensive subjects had no claims for dementia but the medical records indicated that some subjects in our cohort did have a diagnosis of dementia. Therefore, when screening for drug-disease contraindications involving dementia using claims data only, the screen would yield false negatives. Other issues that enter the evaluation of a screening program include four characteristics: the clinical importance of the conditions that are being screened; existence of an efficacious treatment for the problems detected; acceptability of the screening test by patients and providers; and, the ability of the system to handle both the population size and the volume of (true or false) positive screens.

DUR screening consists of applying content validated criteria to a patient's medication history. To allocate resources and run efficient DUR programs, policy makers need to know the sensitivity and specificity of their DUR programs. They should consider resources spent on false positive flags and the public health risk for false negative flags. This research gives information for at least one chronic disease (hypertension) and the sensitivity and specificity of a computerized DUR screening program for its treatment. Neither the sensitivity nor specificity was sufficiently high enough to qualify for an efficient screen of inappropriateness of antihypertensive drug therapy. Programs that employ such algorithms should use caution in denying payment or basing clinical decisions solely on such mechanisms.

Improvement of the application of under-utilization and over-utilization flags may improve the screen's specificity to detect drug therapy inappropriateness. However, we conclude that a highly specific and sensitive screen requires more information than is currently available through administrative claims data. Specifically, clinical markers of drug therapy effectiveness may significantly improve the screen's sensitivity and specificity. Incorporation of clinical data should be feasible, especially for managed care organizations that take advantage of computerized medical records. DUR program managers should encourage development of this technology. Although medical record data are "closer to the source" than administrative claims data, medical record data are far from complete. However, it is from medical record data that medical care decisions are made. Thus, incorporation of medical record information into DUR screening processes represents a more realistic approach to evaluating drug therapy inappropriateness.

The results of this research can be used by state Medicaid program policy makers who are responsible for ensuring drug therapy appropriateness. HCFA should be especially interested in these results since it is the federal agency responsible for ensuring that states comply with OBRA 1990 legislation. Also, state agencies and individuals involved in implementing DUR programs under Medicaid frequently look to HCFA for guidance in designing a DUR program that will have a high likelihood of screening for drug therapy inappropriateness. In addition, HCFA is responsible for overseeing the evaluation of the prospective DUR demonstration project, and recommending policy based on its results. It is imperative that policy makers know the relative sensitivity and specificity of the drug therapy inappropriateness measure being used by mandated state Medicaid programs to screen for individual cases of drug therapy inappropriateness. The manual of operations (Appendix A) has been developed to help those evaluating DUR programs, whether in fee-for-service Medicaid programs or managed care environments. When selecting a DUR vendor, one should assure that there has been an assessment of the program's validity; the manual offers a methodology to do this. Unless policy makers demand quality DUR programs from vendors, the "state of the art" for DUR will not improve, and resources will be wasted on ineffective, inefficient DUR programs.

However, it would be costly and unrealistic to fully assess a DUR program's sensitivity and specificity. Alternatively, we recommend that prospective DUR screens be limited to those that, if violated, could lead to an immediate, identifiable threat to patient health. Use of other screening criteria should be limited to retrospective analyses examining drug prescribing patterns rather than identifying inappropriate drug therapy on individual patients.

Figure 1

Histogram of age (N=738)

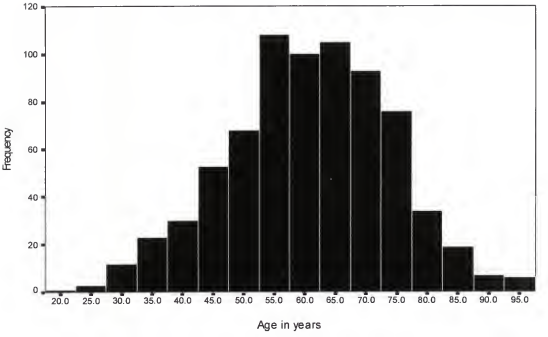


Figure 2

Histogram of number of antihypertensive drugs per subject (N=738)

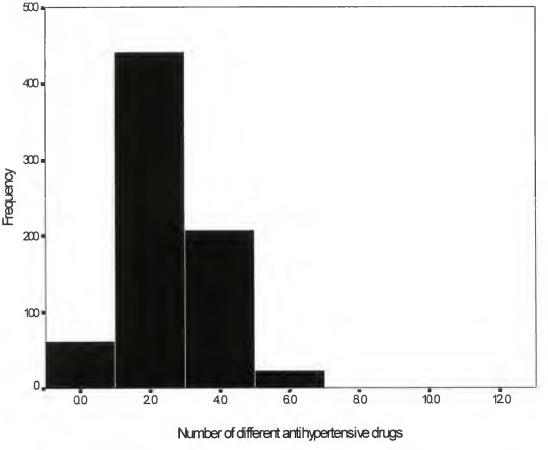


Figure 3

Histogram of number of diagnostic categories per subject (N=738)

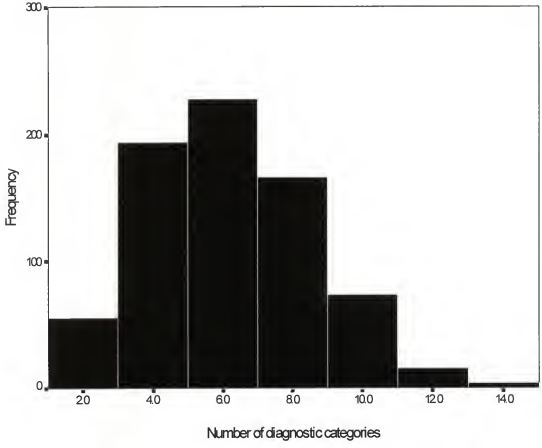


Figure 4

Histogram of compliance ratio (N=631)

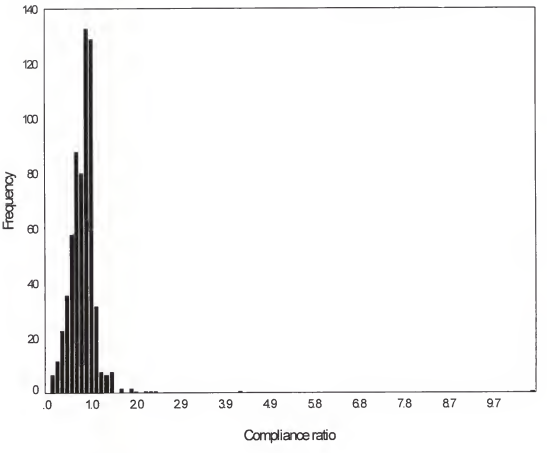
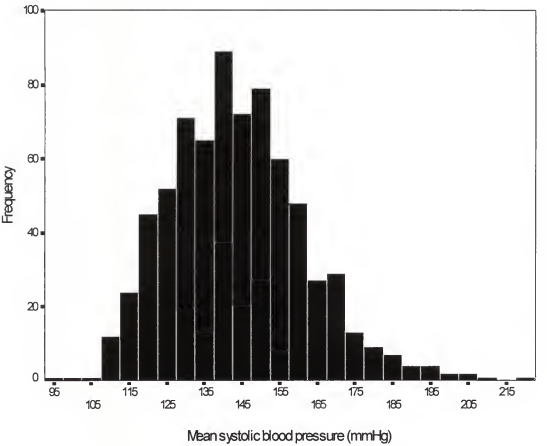


Figure 5

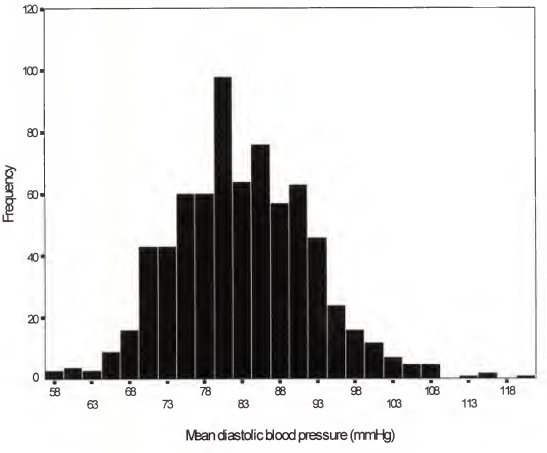
Histogram of mean systolic blood pressure (N=719)



Note: Includes 88 subjects with a single systolic blood pressure reading as the mean.

Figure 6

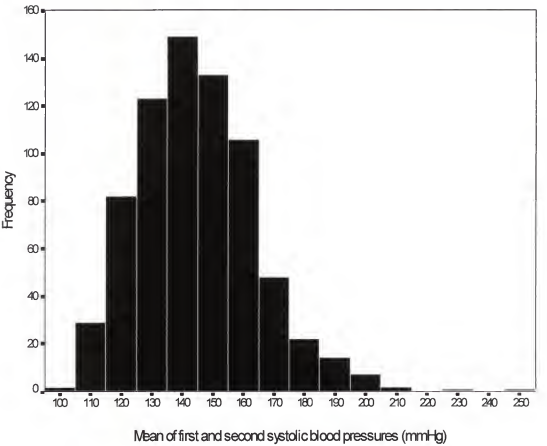
Histogram of mean diastolic blood pressure (N=718)



Note: Includes 88 subjects with a single diastolic blood pressure reading as the mean.

Figure 7

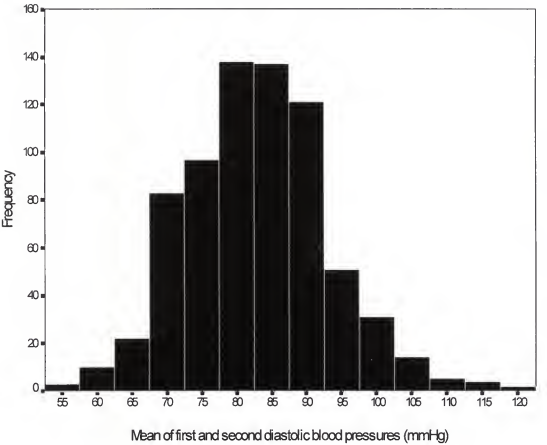
Histogram of mean of first and second systolic blood pressures (N=719)



Note: Includes 88 subjects with a single systolic blood pressure reading as the mean.

Figure 8

Histogram of mean of first and second diastolic blood pressures (N=718)



Note: Includes 88 subjects with a single diastolic blood pressure reading as the mean.

Figure 9

Histogram of percent of uncontrolled systolic blood pressures (N=719)

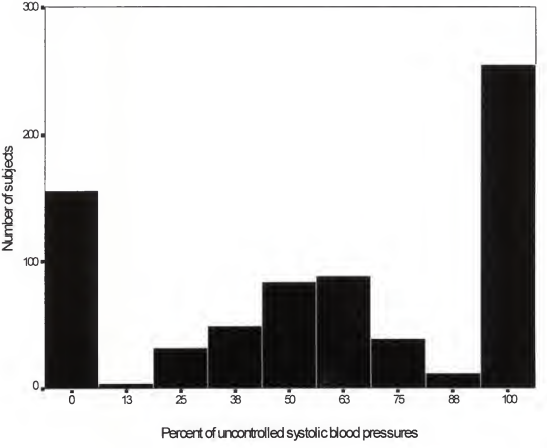


Figure 10

Histogram of percent of uncontrolled diastolic blood pressures (N=718)

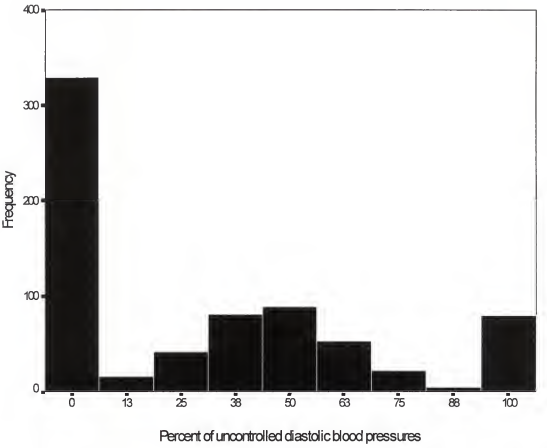
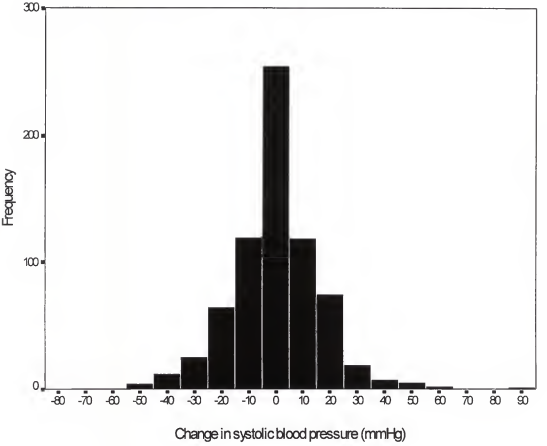


Figure 11

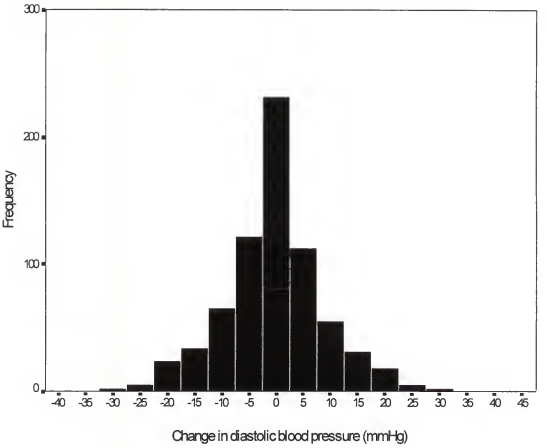
Histogram of change in systolic blood pressure (N=719)



NOTE: "Change in systolic blood pressure" was operationally defined as the last systolic blood pressure reading minus the first systolic blood pressure reading during the study period. Thus, if there were only one blood pressure reading for a subject, then the value is zero.

Figure 12

Histogram of change in diastolic blood pressure (N=718)



NOTE: "Change in diastolic blood pressure" was operationally defined as the last diastolic blood pressure reading minus the first diastolic blood pressure reading during the study period. Thus, if there were only one blood pressure reading for a subject, then the value is zero.

Figure 13

Information flow for DURSCREEN assessment development

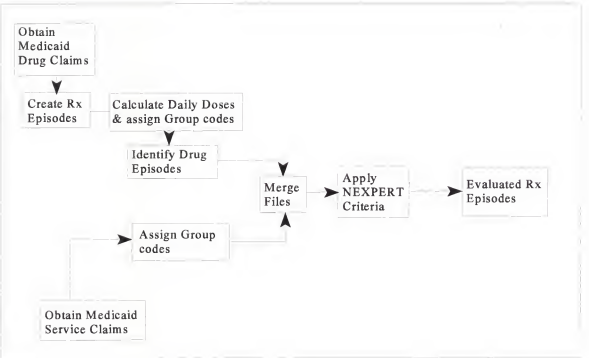


Figure 14

Information flow for INDEPTH assessment development

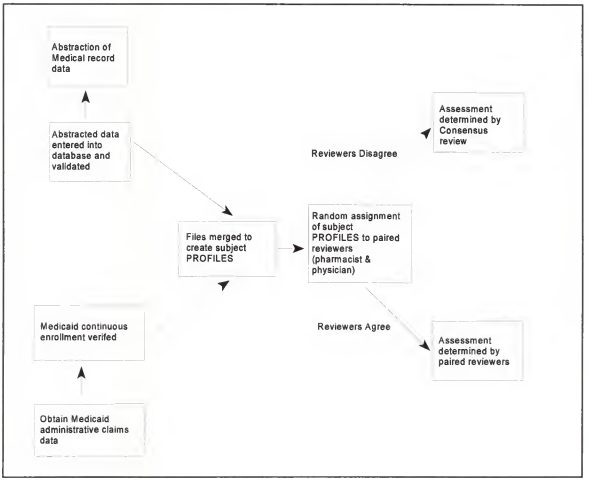
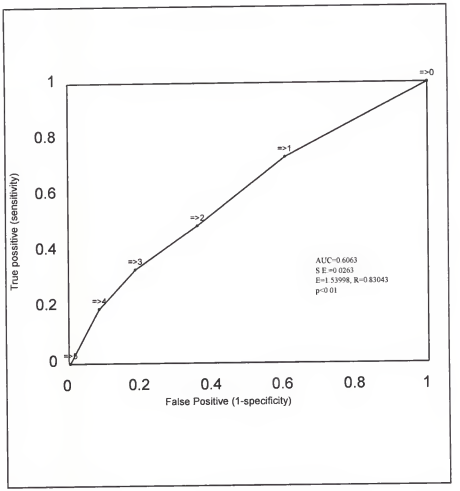
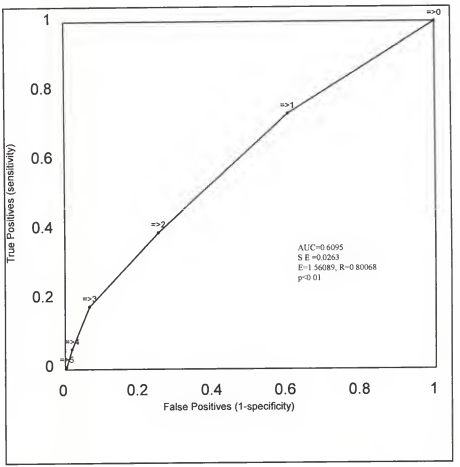


Figure 15
Receiver operating characteristic (ROC) curve for number of flags



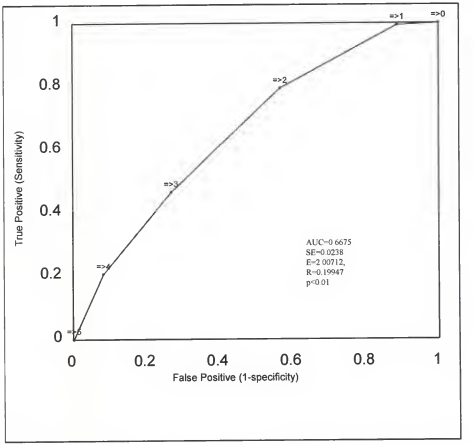
NOTE: The parameter "E" determines the "height" of the curve along the negative diagonal. The parameter "R" determines the "skewness" of the curve with respect to the negative diagonal. AUC is the Area under the curve. "Number of flags" is the sum of the number of flags per subject from the DURSCREEN assessment.

Figure 16
Receiver operating characteristic (ROC) curve for number of criteria element flags



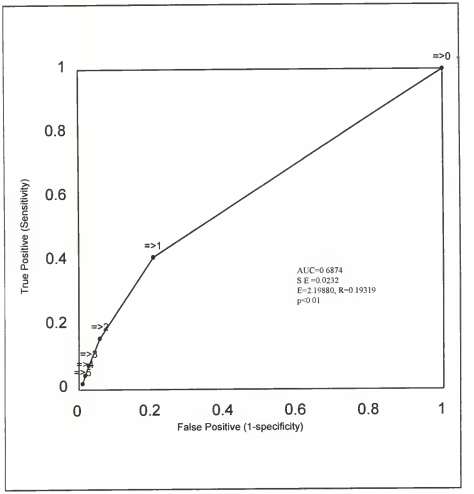
NOTE: The parameter "E" determines the "height" of the curve along the negative diagonal. The parameter "R" determines the "skewness" of the curve with respect to the negative diagonal. AUC is the Area under the curve. A criteria element is a categorization of criteria types and include these six types: therapeutic duplication, incorrect dose, drug-drug interaction, drug-disease contraindication, over-utilization and under-utilization. "Number of criteria element flags" is the sum of the number of elements per subject that were flagged in the DURSCREEN assessment.

Figure 17
Receiver operating characteristic (ROC) curve for number of antihypertensive drugs



NOTE: The parameter "E" determines the "height" of the curve along the negative diagonal. The parameter "R" determines the "skewness" of the curve with respect to the negative diagonal.

Figure 18
Receiver operating characteristic (ROC) curve for number of flags excluding utilization flags [DURSCREEN(4) derivative]



NOTE: The parameter "E" determines the "height" of the curve along the negative diagonal. The parameter "R" determines the "skewness" of the curve with respect to the negative diagonal. DURSCREEN(4) identified those subjects with at least one flag for any criteria but not including a flag for either under-utilization or over-utilization.

Table 1

Percent of subjects, by hospital clinic site

Hospital Clinic Site	Percent of Subjects N=738
Site One	40
Site Two	12
Site Three	12
Site Four	36

Table 2

Descriptive statistics, by select continuous variables

Variable name	Number of subjects[1]	Mean	Median	Standard Deviation	Range
age in years	738	60.6	61.0	13.3	21 to 96
Number of antihypertensive drugs[2]	738	2.0	2	1.3	0 to 11
Number of diagnosis categories[2]	738	5.7	6	2.4	1 to 13
compliance ratio	631	0.84	0.86	0.49	0.12 to 10.52
mean systolic blood pressure (mmHg)	719	143.7	142.3	18.1	94.7 to 221.0
mean diastolic blood pressure (mmHg)	718	82.6	82.0	9.2	57.0 to 120.0
mean of first and second systolic blood pressures (mmHg)	719	144.0	142.0	19.6	96.0 to 247.5
mean of first and second diastolic blood pressures (mmHg)	718	82.8	82.5	10.1	56.0 to 122.0
percent of uncontrolled[3] systolic blood pressures	719	58	67	39	0 to 100
percent of uncontrolled[3] diastolic blood pressures	718	31	20	35	0 to 100
change[4] in systolic blood pressure (mmHg)	719	-0.43	0	17.5	-75 to 85
change[4] in diastolic blood pressure (mmHg)	718	-0.52	0	9.4	-42 to 43

[1]The number of subjects varies due to lack of available data.

[2]While these variables are categorical, because of the range and the normal distribution we elected to treat them as continuous for several of the analyses.

[3]Percent of uncontrolled blood pressure readings - the number of diastolic blood pressures ≥ 90 mmHg or the number of systolic blood pressures ≥ 140 mmHg expressed as a percentage of the total number of diastolic or systolic blood pressure readings per subject, respectively.

[4] Change - the last blood pressure reading minus the first blood pressure reading, if only one blood pressure reading the change value is equal to 0.

Table 3

Study population, by selected demographics:

Demographics		Subjects N=738	
		#	%
Race	Black	658	89
	All Others	80	11
Gender	Female	574	78
	Male	164	22
Age in years (mean + S.D.)		60.6 ± 13.3	

Table 4

Percent of subjects, by information and source

Information	Source	Percent of Subjects N=738
Diagnoses	Chart	100
Diagnoses	Claims	99
Prescription	Claims	99
Laboratories	Chart	99
Procedures	Claims	94
Blood pressure/weight	Chart	97

Table 5

Frequency of subjects' drug use[1],
by the number of different antihypertensive (AHT) drugs

# of different AHT Drugs	Subjects (N=738)	
	#	%
Zero	62	8.4
One	216	29.3
Two	226	30.6
Three	151	20.5
Four	56	7.6
Five	17	2.3
Six	7	0.9
Seven	1	0.1
Ten	1	0.1
Eleven	1	0.1

[1]Use is defined as the presence of a least one prescription claim for an antihypertensive drug during the subjects' nine month study period.

Table 6

Frequency of subjects' use[1] of antihypertensive drugs, by drug class

Antihypertensive Drug Class	Subjects (N=738)	
	#	%
angiotensin converting enzyme inhibitor	223	30.2
beta adrenergic blocking agents	98	13.3
calcium channel blockers	446	60.4
diuretics	398	53.9
alpha-1-adrenergic agents	40	5.4
centrally acting alpha-adrenergic agents	76	10.3
peripherally acting alpha-adrenergic agents	1	0.1
vasodilators	15	2.0

[1]Use is defined as the presence of a least one prescription claim within a drug class during the subject's nine month study period.

Table 7

Definitions for drug use review screening criteria elements

CRITERIA ELEMENT	DEFINITION
ADVERSE DRUG-DRUG INTERACTION	The potential for an adverse medical effect as a result of a person using two or more drugs together.
DRUG-DISEASE CONTRAINDICATION	The potential for an undesirable alteration of the therapeutic effect of a given prescription because of the presence of a disease condition.
INCORRECT DRUG DOSAGE	A dosage that lies outside the daily dosage range specified in expert-developed criteria as necessary to achieve therapeutic benefit.
OVERUTILIZATION	Use of a drug in quantities or for durations that put the recipient at risk of an adverse medical result.
PREGNANCY CONFLICT	Use of a prescribed drug that is not recommended during pregnancy.
THERAPEUTIC DUPLICATION	The prescribing and dispensing of two or more drugs from the same therapeutic class such that the combined daily dose puts the patient at risk of an adverse medical effect
UNDERUTILIZATION	Use of a drug in insufficient quantity to achieve a desired therapeutic goal.

Table 8

Characteristics of Delphi survey participants

Name	Address	Type of Expertise
Emmanuel L. Bravo, M.D.	Cleveland Clinic Foundation 9500 Euclid Avenue Cleveland, OH 44195	as an authority in the treatment of hypertension
Henry I. Bussey, Pharm.D., FCCP	Associate Professor University of Texas at Austin College of Pharmacy Austin, TX 78712-1074	as an authority in the treatment of hypertension
Barry L. Carter, Pharm.D.	Associate Professor and Assistant Head for Ambulatory Care Department of Pharmacy Practice University of Illinois at Chicago College of Pharmacy 833 South Wood Street Chicago, IL 60612-7230	as an authority in the treatment of hypertension
Ray W. Gifford, Jr., M.D.	Vice Chairman, Division of Medicine Cleveland Clinic Foundation 9500 Euclid Avenue Cleveland, OH 44195	as an authority in the treatment of hypertension
Barry M. Massie, M.D.	Director, Coronary Care Unit & Hypertension Department of Medicine, Cardiology Section San Francisco Veterans Affairs Medical Center Room 111C 4150 Clement San Francisco, 94121	as an authority in the treatment of hypertension
Robert J. Michocki, Pharm.D.	Professor and Clinical Associate Professor Family Medicine University of Maryland at Baltimore Schools of Pharmacy and Medicine 100 Penn Street, Room 205 Baltimore, MD 21201-1180	as an authority in the treatment of hypertension
Peter Rudd, M.D.	Professor of Medicine Stanford University Clinics Stanford University Medical Center Room X-216 MSOB Stanford, CA 94305-5475	as an authority in the treatment of hypertension

Name	Address	Type of Expertise
Robert L. Talbert, Pharm.D.	Professor University of Texas at Austin College of Pharmacy Austin, TX 78712-1074	as an authority in the treatment of hypertension
William B. White, M.D.	Professor of Medicine Chief, Section of Hypertension and Vascular Diseases University of Connecticut Health Center 263 Farmington Avenue Farmington, CT 06030-3940	as an authority in the treatment of hypertension
Sheldon G. Sheps, M.D.	Chair, Division of Hypertension and Internal Medicine Mayo Medical School and Clinic 200 - 1st Avenue SW Rochester, MN 55905-0001	as an epidemiologist with published studies in the field of hypertension
Fran C. Wheeler, Ph.D.	Director, Center for Health Promotion South Carolina Department of Health & Environmental Control 2600 Bull Street Columbia, SC 29201	as an epidemiologist with published studies in the field of hypertension
Aram V. Chobanian, M.D.	Dean and Professor of Medicine Boston University School of Medicine 80 E. Concord Street L-103 Boston, MA 02118	as a researcher with expertise in the treatment of hypertension
Edward D. Frohlich, M.D.	Vice President for Academic Affairs Alton Ochsner Medical Foundation 1516 Jefferson Highway New Orleans, LA 70121-2484	as a researcher with expertise in the treatment of hypertension
Norman M. Kaplan, M.D.	Professor of Internal Medicine University of Texas Southwestern Medical School 5323 Harry Hines Boulevard Dallas, TX 75235	as a researcher with expertise in the treatment of hypertension
Jackson T. Wright, Jr., M.D., Ph.D.	Division of Hypertension, School of Medicine Case Western Reserve University Room W165 10900 Euclid Avenue Cleveland, OH 44106-4982	as a researcher with expertise in the treatment of hypertension

Table 9

Delphi criteria acceptance rates, by drug class

Drug class	Number of Initial Criteria	Number of Final Criteria (with or without modification)	Acceptance Rate
Angiotensin converting enzyme inhibitors	13	10	0.77
Calcium Channel Blockers	15	14	0.93
Beta-blockers	22	15	0.68
Diuretics	24	23	0.96
Other Antihypertensives	40	30	0.75
TOTAL	114	92	0.81

Table 10

Subjects, by DURSCREEN assessment

Assessment Classification	Number of Subjects	Percent of Subjects
Appropriate	272	36.9
Not Appropriate	466	63.1
TOTAL	738	100.0

Table 11

Subjects identified by DURSCREEN as inappropriate[1], by criteria element

Criteria Element	Number of subjects	% of Subjects
Dose	84	11.4
Duplication	5	0.7
Drug-disease	38	5.1
Drug-drug Interaction	114	15.4
Over-utilization	170	23.0
Under-utilization	345	46.7

NOTE: Subjects may be included in more than one category.

[1]Inappropriate is defined as having at least one flag within the element category.

Table 12
Subjects' flag frequency by specific criteria

CRITERIA		Subjects' flag frequency										Subjects > 1 Flags	% of Subjects > 1 Flag
Identification Number and Description		*0 Flags*		*1 Flag*		*2 Flags*		*3 Flags*		*4 Flags*			
1	Dose (ALL Drugs)	654	88.6%	67	9.1%	14	1.90%	-	0.0%	-	0.00%	84	11.4%
2	Duplication (ALL Drugs)	733	99.3%	5	0.7%	-	0.0%	-	0.0%	-	0.0%	5	0.7%
3	Accessory bypass tract condition & diltiazem/verapamil	709	100.0%	-	0.0%	3	0.0%	-	0.0%	-	0.0%	0	0.0%
4	Anuria and Diuretics	737	99.5%	1	0.1%	-	0.0%	-	0.0%	-	0.0%	1	0.1%
5	Bronchospasm/Asthma and beta-blockers	737	99.9%	1	0.1%	-	0.0%	-	0.0%	-	0.0%	1	0.1%
6	Congestive heart failure and beta-blockers or calcium channel blockers	709	96.1%	25	3.4%	3	0.4%	-	0.1%	-	0.0%	29	3.9%
7	CAD and hydralazine or minoxidil	735	99.7%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	2	0.0%
8	Dementia & centrally acting adrenergic agents or reserpine	709	100.0%	-	0.0%	3	0.0%	-	0.0%	-	0.0%	0	0.0%
9	Diabetic nephropathy & potassium sparing diuretics	709	100.0%	-	0.0%	3	0.0%	-	0.0%	-	0.0%	0	0.0%
10	Hepatic impairment & methyldopa or verapamil	709	100.0%	-	0.0%	3	0.0%	-	0.0%	-	0.0%	0	0.0%
11	Depression and reserpine or rauwolfia	738	100.0%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	0	0.0%
12	Hyperkalemia & angiotensin converting enzyme inhibitors	738	100.0%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	0	0.0%
13	Peptic ulcer disease & peripherally acting adrenergic agents	738	100.0%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	0	0.0%
14	Pericardial effusion & minoxidil	738	100.0%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	0	0.0%

CRITERIA		Subjects' flag frequency										Subjects > 1 Flags	% of Subjects > 1 Flag
Identification Number and Description		*0 Flags*		*1 Flag*		*2 Flags*		*3 Flags*		*4 Flags*			
15	Peripheral vascular disease and non-selective beta-blockers	736	99.7%	2	0.3%	-	0.0%	-	0.0%	-	0.0%	2	0.3%
16	Pheochromocytoma & guanethidine/guanadrel/minoxidil	738	100.0%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	0	0.0%
17	Pregnancy & select antihypertensives	738	100.0%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	0	0.0%
18	Renal artery stenosis & angiotensin converting enzyme inhibitors	738	100.0%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	0	0.0%
19	Raynaud's disease & non-selective beta-blockers	735	100.0%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	0	0.0%
20	Renal impairment & potassium sparing diuretics	735	100.0%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	0	0.0%
21	Heart block & beta-blockers	738	100.0%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	3	0.0%
22	Sick sinus syndrome and diltiazem/verapamil	735	99.6%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	3	0.4%
23	Ulcerative colitis & reserpine/rauwolfia	738	100.0%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	0	0.0%
24	Potassium sparing diuretic and angiotensin converting enzyme inhibitors	723	98.0%	-	0.0%	14	0.0%	-	0.0%	-	0.0%	15	2.0%
25	Verapamil and carbamazepine	735	99.6%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	3	0.0%
26	Cardiac glycoside and verapamil/potassium wasting diuretic without potassium supplement or angiotensin converting enzyme inhibitor	726	98.4%	12	1.6%	-	0.0%	-	0.0%	-	0.0%	12	1.6%
27	Aminophylline & beta-blockers	738	100.0%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	0	0.0%
28	Amphetamines & guanethidine/guanadrel	738	100.0%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	0	0.0%

CRITERIA		Subjects' flag frequency										Subjects > 1 Flags	% of Subjects > 1 Flag
Identification Number and Description		*0 Flags*		*1 Flag*		*2 Flags*		*3 Flags*		*4 Flags*			
29	Amphotericin B & potassium wasting diuretics	738	100.0%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	0	0.0%
30	Bupropion & methyldopa/guanethidine/guanadrel	738	100.0%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	0	0.0%
31	Cholestyramine/colestipol and potassium wasting diuretics	734	99.5%	4	0.5%	-	0.0%	-	0.0%	-	0.0%	4	0.5%
32	Corticosteroids and potassium wasting diuretics	721	97.7%	17	2.3%	-	0.0%	-	0.0%	-	0.0%	17	2.3%
33	Cyclosporine & diltiazem/verapamil/potassium sparing diuretics	738	100.0%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	0	0.0%
34	Ergot alkaloids and beta-blockers	737	99.9%	-	0.1%	-	0.0%	-	0.0%	-	0.0%	1	0.1%
35	Haloperidol & guanethidine/guanadrel	738	100.0%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	0	0.0%
36	Indomethacin and diuretics	727	98.5%	9	1.2%	2	0.3%	-	0.0%	-	0.0%	11	1.5%
37	Angiotensin converting enzyme inhibitors and potassium sparing diuretics	723	98.0%	15	2.0%	-	0.0%	-	0.0%	-	0.0%	15	2.0%
38	Potassium supplement and angiotensin converting enzyme inhibitors or potassium sparing diuretic without a potassium wasting diuretic	711	96.3%	27	3.7%	-	0.0%	-	0.0%	-	0.0%	27	3.7%
39	Levodopa and centrally acting adrenergic agents/ reserpine/rauwolfia	738	100.0%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	0	0.0%
40	Lithium & angiotensin converting enzyme inhibitor or potassium wasting diuretics	735	99.6%	2	0.3%	1	0.1%	-	0.0%	-	0.0%	3	0.4%

CRITERIA		Subjects' flag frequency										Subjects ≥ 1 Flags	% of Subjects ≥ 1 Flag
Identification Number and Description		*0 Flags*		*1 Flag*		*2 Flags*		*3 Flags*		*4 Flags*			
41	Monoamine oxidase inhibitors & beta-blockers or peripherally acting adrenergic agents	738	100.0%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	0	0.0%
42	Maprotiline & methyldopa/guanethidine/guanadrel	738	100.0%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	0	0.0%
43	Methylphenidate & guanethidine/guanadrel	738	100.0%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	0	0.0%
44	Oral anticoagulants & ethacrynic acid	738	100.0%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	0	0.0%
45	Phenothiazines & guanethidine/guanadrel	738	100.0%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	0	0.0%
46	Salicylates & furosemide	738	100.0%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	0	0.0%
47	Sympathomimetics & non-selective beta-blockers or methyldopa	738	100.0%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	0	0.0%
48	Theophylline and beta-blockers	735	99.7%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	0	0.3%
49	Thioxanthines and guanethidine/guanadrel	738	100.0%	-	0.0%	-	0.0%	-	0.0%	-	0.0%	0	0.0%
50	Trazodone and adrenergic agents	737	99.9%	1	0.1%	-	0.0%	-	0.0%	-	0.0%	1	0.1%
51	Tricyclic antidepressants and adrenergic agents	735	99.6%	3	0.4%	-	0.0%	-	0.0%	-	0.0%	3	0.4%
52	Over-utilization (ALL Drugs)	568	77.0%	118	16.0%	45	6.1%	6	0.8%	1	0.1%	170	23.0%
53	Under-utilization (ALL Drugs)	393	53.3%	200	27.1%	114	15.4%	27	3.7%	4	0.5%	345	46.7%

Table 13

Subjects' DURSCREEN assessment, by number of flags

FLAGS	DURSCREEN APPROPRIATE		DURSCREEN INAPPROPRIATE		ROW SUMS	SUBJECTS
#	#	%	#	%	#	%
0	272	100	0	0	272	37
1	0	0	189	41	189	26
2	0	0	119	26	119	16
3	0	0	78	17	78	11
4	0	0	39	8	39	5
>5	0	0	41	9	41	6
TOTAL	272	100	466	100	738	100

Table 14

Subjects' DURSCREEN assessment excluding utilization[1], by flag frequency

FLAGS	DURSCREEN APPROPRIATE		DURSCREEN INAPPROPRIATE		ROW SUMS	SUBJECTS
#	#	%	#	%	#	%
0	272	100.0	0	0.0	272	59.9
1	0	0.0	123	67.6	123	27.1
2	0	0.0	30	16.5	30	6.6
3	0	0.0	10	5.5	10	2.2
4	0	0.0	9	4.9	9	2.0
>5	0	0.0	10	5.5	10	2.2
TOTAL	272	100.0	182	100.0	454	100.0

[1]Those subjects who received only a flag for the under-utilization or over-utilization criteria are excluded (N=284).

Table 15

Subjects' DURSCREEN assessment, by the frequency of unique criteria flags

# of unique criteria flags	DURSCREEN APPROPRIATE		DURSCREEN INAPPROPRIATE		ROW SUMS	SUBJECTS
#	#	%	#	%	#	%
0	272	100.0	0	0.0	272	36.9
1	0	0.0	265	56.9	265	35.9
2	0	0.0	136	29.2	136	18.4
3	0	0.0	45	9.7	45	6.1
4	0	0.0	16	3.4	16	2.2
>5	0	0.0	4	0.9	4	0.5
TOTAL	272	100.0	466	100.0	738	100.0

NOTE: The total number of possible criteria flags are 53.

Table 16

Subjects' INDEPTH assessment, by paired individual reviewer assessments (physician and pharmacist) (N=738)

Paired Individual Reviewer Assessments		INDEPTH Assessment			
		Appropriate	Inappropriate	Cannot Determine	Row sums
PHYSICIAN	PHARMACIST	Number (column %)	Number (column %)	Number (column %)	Number (column %)
Appropriate	Appropriate	375 (77.6)	...	...	375 (50.8)
Appropriate	Inappropriate	50 (10.4)	34 (21.9)	15 (15,0)	...
Appropriate	Cannot Determine	20 (4.1)	5 (3.2)	1ů (11.0)	...
Inappropriate	Appropriate	17 (3.5)	22 (14.2)	11 (11.0)	...
Inappropriate	Inappropriate	...	72 (46.5)	...	72 (9.8)
Inappropriate	Cannot Determine	1 (0.2)	8 (5.2)	8 (8.0)	...
Cannot Determine	Appropriate	14 (2.9)	4 (2.6)	8 (8.0)	...
Cannot Determine	Inappropriate	6 (1.2)	10 (6.5)	15 (15.0)	...
Cannot Determine	Cannot Determine	...	...	33 (33.0)	33 (4.5)
TOTALS		483 (100)	155 (100)	100 (100)	738

Note: If the paired reviewers' individual assessments matched, the INDEPTH assessment was preserved. If, the reviewers' individual assessments did not match, the INDEPTH assessment was determined by the consensus panel.

Table 17

Subjects' INDEPTH assessment, by diagnostic groupings

Diagnostic Group (ICD-9 Code)	Inappropriate # (%)[1]	Appropriate # (%)[1]
Hypertension and complications (401-404)	155 (100)	483 (100)
Other Circulatory Diseases (390-400; 405-459)	84 (54.2)	246 (50.9)
Endocrine and Immune (240-279)	108 (69.7)	337 (69.8)
Symptoms, Signs and Ill-defined Conditions (780-799)	89 (57.4)	287 (59.4)
Musculoskeletal System and Connective Tissue (710-739)	76 (49.0)	261 (54.0)
Genitourinary (580-629)	69 (44.5)	204 (42.2)
Nervous System and Sense Organs (320-389)	61 (39.4)	208 (43.1)
Mental[†] (290-319)	47 (30.3)	100 (41.2)
Respiratory (460-519)	68 (43.9)	173 (35.8)
Digestive (520-579)	52 (33.5)	191 (39.5)
Infections and Parasitic (1-139)	44 (28.4)	104 (21.5)
Blood and Blood-forming Organs (280-289)	33 (21.3)	93 (19.3)
Neoplasm (140-239)	23 (14.8)	87 (18.0)
Injury and Poisoning (800-999)	32 (20.6)	82 (17.0)
Skin and Subcutaneous Tissue (680-709)	25 (16.1)	65 (13.5)
Complications of Pregnancy and Childbirth (630-676 & 740-779)	13 (8.4)	32 (6.6)

[†]Significant $p<0.05$

[1]The reported percents are determined by the number of subjects who had a particular disease category (claims and/or medical chart) and the total of number of subjects identified by the Indepth assessment as "appropriate" (N=483) or "inappropriate" (N=155).

NOTE: One way analysis of variance was used to assess for statistically significant differences.

Table 18

Mean blood pressure readings, by INDEPTH assessment

INDEPTH Assessment	Mean (+ SE) of all Diastolic BP Readings	Mean (+ SE) of all Systolic BP Readings
Inappropriate (N = 153)	88.43 (+0.86)	159.68 (+1.61)
Appropriate (N = 483)	79.88 (+0.32)	137.54 (+0.62)
	F = 126.5; p = 0.00	F = 240.8; p = 0.00

NOTE: N=636 subjects. Two subjects did not have any blood pressure readings and therefore were excluded from this analysis. One way analysis of variance was used to assess for statistically significant differences.

Table 19

Mean of the 1st and 2nd blood pressure readings, by INDEPTH assessment

INDEPTH Assessment	Mean (+ SE) of 1st & 2nd Diastolic BP Readings	Mean (+ SE) of 1st and 2nd Systolic BP Readings
Inappropriate (N = 153)	88.38 (+0.93)	159.01 (+1.75)
Appropriate (N = 483)	80.38 (+0.39)	138.31 (+0.73)
	F = 84.0; p = 0.00	F = 161.2; p = 0.00

NOTE: N=636 subjects. Two subjects did not have any blood pressure readings and therefore were excluded from this analysis. One way analysis of variance was used to assess for statistically significant differences.

Table 20

Mean change in blood pressure readings by INDEPTH assessment

INDEPTH Assessment	Mean Change (+ SE) in Diastolic BP	Mean Change (+ SE) in Systolic BP
Inappropriate (N = 153)	0.75 (+0.90)	2.78 (+1.69)
Appropriate (N = 483)	-1.27 (+0.40)	-2.01 (+0.72)
	$Z = 1.3$; $p < 0.01$	$Z = 1.8$; $p = 0.00$

NOTE: N=636 subjects. Two subjects did not have any blood pressure readings and therefore were excluded from this analysis. The Kolmogorov-Smirnov test was used to assess for statistically significant differences.

Table 21

Mean percent of uncontrolled blood pressure readings, by INDEPTH assessment

INDEPTH Assessment	Mean % (+ SE) of Uncontrolled Diastolic BP Readings (> 90 mmHg)	Mean % (+ SE) of Uncontrolled Systolic BP Readings (> 140 mmHg)
Inappropriate (N = 153)	53 (+3.1)	82 (+2.3)
Appropriate (N = 483)	19 (+1.2)	48 (+1.7)
	Z = 4.4; p = 0.00	Z = 4.5; p = 0.00

NOTE: N=636 subjects. Two subjects did not have any blood pressure readings and therefore were excluded from this analysis. The Kolmogorov-Smirnov test was used to assess for statistically significant differences.

Table 22

Comparison of INDEPTH assessment, by DURSCREEN assessment

DURSCREEN Assessment	INDEPTH Assessment		
	INAPPROPRIATE	APPROPRIATE	ROW SUMS
INAPPROPRIATE	114	292	406
APPROPRIATE	41	191	232
COLUMN TOTALS	155	483	638

NOTE: Sensitivity = 73.5%; Specificity = 39.5%; Percent Agreement = 47.9%.

Table 23

Evaluation of Sensitivity and Specificity of DURSCREEN and DURSCREEN derivatives (N=638)

DURSCREEN derivative	Definition of INAPPROPRIATENESS	Number of true positives (N=155)	Number of true negatives (N=483)	*p*-value	Sensitivity	Specificity	Percent Agreement
DURSCREEN	at least one flag	114	191	0.00319	0.735	0.395	47.8
DURSCREEN(2)	at least one OVERUTILIZATION flag	95	384	0.00000	0.297	0.785	66.6
DURSCREEN(3)	at least one UNDERUTILIZATION flag	87	256	0.56114	0.497	0.53	52.2
DURSCREEN(4)	at least one flag except for ALL UTILIZATION	64	384	0.00000	0.413	0.795	70.2
DURSCREEN(5)	at least one flag except for UNDERUTILIZATION	87	308	0.00000	0.561	0.643	61.9
DURSCREEN(6)	at least one flag except for OVERUTILIZATION	101	224	0.01171	0.652	0.464	50.9
DURSCREEN(7)	at least one OVER OR UNDERUTILIZATION flag	95	214	0.22066	0.613	0.443	48.4
DURSCREEN(8)	at least one DOSE flag	33	436	0.00016	0.213	0.903	73.5
DURSCREEN(9)	at least one DRUG-DRUG INTERACTION flag	31	433	0.00169	0.200	0.896	72.7
DURSCREEN(10)	at least one DRUG-DISEASE CONTRAINDICATION flag	44	423	0.00000	0.284	0.876	73.2

Table 24

Receiver operating characteristic curve data table for DURSCREEN, by number of flags

Cutoff	True Positive (Sensitivity)	False Positive (1-Specificity)
> 0 flags	1.000	1.000
> 1 Flag	0.736	0.605
≥ 2 Flags	0.490	0.360
≥ 3 Flags	0.336	0.186
> 4 Flags	0.194	0.085
≥ 5 Flags	0.116	0.037

NOTE: Data in this table correspond to Figure 15.

Table 25

Receiver operating characteristic curve data table for DURSCREEN, by the number of criteria element flags

Cutoff	True Positive (Sensitivity)	False Positive (1-Specificity)
> 0 Flag Elements	1.000	1.000
> 1 Flag Elements	0.736	0.605
> 2 Flag Elements	0.394	0.253
> 3 Flag Elements	0.181	0.080
> 4 Flag Elements	0.058	0.019
> 5 Flag Elements	0.007	0.004

NOTE: Data in this table correspond to Figure 16.

Table 26

Receiver operating characteristic curve data table for DURSCREEN, by the number of antihypertensive drugs

Cutoff	True Positive (Sensitivity)	False Positive (1-Specificity)
≥ 0 Antihypertensive drugs	1.000	1.000
> 1 Antihypertensive drugs	0.994	0.882
> 2 Antihypertensive drugs	0.794	0.569
> 3 Antihypertensive drugs	0.500	0.269
≥ 4 Antihypertensive drugs	0.207	0.000
> 5 Antihypertensive drugs	0.071	0.023

NOTE: Data in this table correspond to Figure 17.

Table 27

Receiver operating characteristic curve data table for DURSCREEN(4) derivative (excluding utilization flags), by number of flags

Cutoff	True Positive (Sensitivity)	False Positive (1-Specificity)
≥ 0 Flags	1.000	1.000
≥ 1 Flags	0.413	0.205
≥ 2 Flags	0.161	0.058
≥ 3 Flags	0.077	0.027
≥ 4 Flags	0.045	0.016
≥ 5 Flags	0.019	0.010

NOTE: Data in this table correspond to Figure 18.

TABLE 28

Model and variable significance for the mean of the 1st and 2nd systolic blood pressure readings, by DURSCREEN and derivatives

	MODEL			VARIABLES				
DURSCREEN derivative	MULTIPLE R	ADJUSTED R^2	F (p- value)	DURSCREEN derivative p -value	AGE p -value	Compliance RATIO p -value	# of AHT drugs p-value	# of Diagnoses Categories p-value
DURSCREEN	0.32	0.09	12.17 (0.00)	0.57	0.00	0.76	0.00	0.067
DURSCREEN(2)	0.32	0.09	12.30 (0.00)	0.35	0.00	1.00	0.00	0.067
DURSCREEN (3)	0.32	0.09	12.23 (0.00)	0.44	0.00	0.80	0.00	0.067
DURSCREEN (4)	0.32	0.09	12.16 (0.00)	0.61	0.00	0.82	0.00	0.072
DURSCREEN (5)	0.32	0.09	12.10 (0.00)	0.96	0.00	0.76	0.00	0.062
DURSCREEN (6)	0.32	0.09	12.21 (0.00)	0.44	0.00	0.87	0.00	0.067
DURSCREEN (7)	0.32	0.09	12.13 (0.00)	0.44	0.00	0.76	0.00	0.066
DURSCREEN (8)	0.32	0.09	12.35 (0.00)	0.29	0.00	0.66	0.00	0.050
DURSCREEN(9)	0.32	0.09	12.33 (0.00)	0.31	0.00	0.79	0.00	0.076
DURSCREEN(10)	0.32	0.09	12.18 (0.00)	0.54	0.00	0.71	0.00	0.051
INDEPTH ASSESSMENT	0.51	0.25	37.08 (0.00)	0.00	0.00	0.16	0.00	0.028

TABLE 29

Model and variable significance for the mean systolic blood pressure readings, by DURSCREEN and derivatives

	MODEL			VARIABLES				
DURSCREEN derivative	MULTIPLE R	ADJUSTED R^2	F (*p*-value)	DURSCREEN DERIVATIVE *p*-value	AGE *p*-value	Compliance RATIO *p*-value	# of AHT drugs *p*-value	# of Diagnoses Categories *p*-value
DURSCREEN	0.33	0.10	12.90 (0.00)	0.97	0.00	0.78	0.00	0.09
DURSCREEN(2)	0.33	0.10	13.24 (0.00)	0.22	0.00	0.92	0.00	0.07
DURSCREEN (3)	0.33	0.10	12.90 (0.00)	0.88	0.00	0.80	0.00	0.09
DURSCREEN (4)	0.33	0.10	12.90 (0.00)	0.91	0.00	0.767	0.00	0.09
DURSCREEN (5)	0.33	0.10	13.00 (0.00)	0.51	0.00	0.84	0.00	0.07
DURSCREEN (6)	0.33	0.10	12.91 (0.00)	0.97	0.00	0.92	0.00	0.09
DURSCREEN (7)	0.33	0.10	12.90 (0.00)	0.88	0.00	0.78	0.00	0.09
DURSCREEN (8)	0.33	0.10	13.11 (0.00)	0.33	0.00	0.68	0.00	0.07
DURSCREEN(9)	0.33	0.10	12.90 (0.00)	0.93	0.00	0.78	0.00	0.09
DURSCREEN(10)	0.33	0.10	13.06 (0.00)	0.39	0.00	0.70	0.00	0.07
INDEPTH ASSESSMENT	0.58	0.33	55.96 (0.00)	0.00	0.00	0.08	0.02	0.03

Table 30

Model and variable significance for the mean of the 1st and 2nd diastolic blood pressure readings, by DURSCREEN and derivatives

	MODEL			VARIABLES				
DURSCREEN derivative	MULTIPLE R	ADJUSTED R^2	F (*p*-value)	DURSCREEN DERIVATIVE *p*-value	AGE *p*-value	Compliance RATIO *p*-value	# of AHT drugs *p*-value	# of Diagnoses Categories *p*-value
DURSCREEN	0.22	0.04	5.50 (0.00)	0.18	0.00	0.11	0.19	0.22
DURSCREEN(2)	0.21	0.03	4.84 (0.00)	0.18	0.00	0.17	0.33	0.17
DURSCREEN (3)	0.22	0.04	5.40 (0.00)	0.16	0.00	0.09	0.25	0.21
DURSCREEN (4)	0.21	0.04	5.22 (0.00)	0.18	0.00	0.11	0.19	0.24
DURSCREEN (5)	0.22	0.04	5.28 (0.00)	0.15	0.00	0.21	0.19	0.24
DURSCREEN (6)	0.22	0.04	5.40 (0.00)	0.16	0.00	0.14	0.25	0.21
DURSCREEN (7)	0.22	0.04	5.27 (0.00)	0.15	0.00	0.09	0.24	0.22
DURSCREEN (8)	0.21	0.04	5.05 (0.00)	0.32	0.00	0.12	0.25	0.19
DURSCREEN(9)	0.21	0.04	5.20 (0.00)	0.19	0.00	0.14	0.21	0.22
DURSCREEN(10)	0.21	0.04	4.94 (0.00)	0.48	0.00	0.13	0.26	0.20
INDEPTH ASSESSMENT	0.39	0.15	19.55 (0.00)	0.00	0.00	0.51	0.48	0.12

TABLE 31

Model and variable significance for diastolic mean blood pressure readings, by DURSCREEN and derivatives

	MODEL			VARIABLES				
DURSCREEN derivative	MULTIPLE R	ADJUSTED R^2	F (*p*-value)	DURSCREEN DERIVATIVE *p*-value	AGE *p*-value	Compliance RATIO *p*-value	# of AHT drugs *p*-value	# of Diagnoses Categories *p*-value
DURSCREEN	0.24	0.05	6.61 (0.00)	0.37	0.00	0.02	0.05	0.16
DURSCREEN(2)	0.24	0.05	6.52 (0.00)	0.57	0.00	0.02	0.08	0.11
DURSCREEN (3)	0.24	0.05	6.86 (0.00)	0.16	0.00	0.01	0.05	0.16
DURSCREEN (4)	0.24	0.05	6.73 (0.00)	0.25	0.00	0.02	0.04	0.17
DURSCREEN (5)	0.24	0.05	6.61 (0.00)	0.39	0.00	0.03	0.04	0.16
DURSCREEN (6)	0.24	0.05	6.81 (0.00)	0.45	0.00	0.01	0.04	0.15
DURSCREEN (7)	0.24	0.05	6.57 (0.00)	0.45	0.00	0.02	0.05	0.16
DURSCREEN (8)	0.24	0.05	6.87 (0.00)	0.16	0.00	0.01	0.04	0.15
DURSCREEN (9)	0.24	0.05	6.48 (0.00)	0.68	0.00	0.02	0.06	0.14
DURSCREEN (10)	0.24	0.05	6.77 (0.00)	0.21	0.00	0.02	0.04	0.18
INDEPTH ASSESSMENT	0.46	0.20	29.17 (0.00)	0.00	0.00	0.16	0.88	0.07

Table 32

Regression models for systolic and diastolic blood pressure
DOSE criterion

	Mean systolic blood pressure		Mean diastolic blood pressure	
Parameters	Beta Coefficients (S.E.)			
Constant	122.89 †	(4.97)	97.21 †	(2.51)
Dose (criterion #1)	1.49	(2.08)	-1.62	(1.04)
Age	0.30 †	(.06)	-0.17 †	(.03)
# of Diagnostic categories	-0.53	(.31)	-0.27	(.16)
Compliance ratio	0.09	(2.34)	-4.00 †	(1.17)
# of Antihypertensive drugs	2.26 †	(.60)	0.49	(.30)
Model Statistics				
R^2	0.09		0.07	
Adjusted R^2	0.08		0.06	
F (d.f.)	12.11 (609)		9.42 (608)	
p-value	0.00		0.00	

†$p<0.05$

Table 33

Regression models for systolic and diastolic blood pressure
DUPLICATION criterion

	Mean systolic blood pressure		Mean diastolic blood pressure	
Parameters	Beta Coefficients (S.E.)			
Constant	122.62 †	(4.90)	96.75 †	(2.50)
Duplication (criterion #2)	-28.22 †	(7.80)	-6.91	(3.94)
Age	0.31 †	(.06)	-0.17 †	(.03)
# of Diagnostic categories	-0.51	(.31)	-0.29	(.16)
Compliance ratio	-0.63	(2.30)	-3.88 †	(1.16)
# of Antihypertensive drugs	2.72 †	(.60)	0.51	(.30)
Model Statistics				
R^2	0.11		7.00	
Adjusted R^2	0.10		0.07	
F (d.f.)	14.96 (609)		9.56 (608)	
p-value	0.00		0.00	

†$p<0.05$

Table 34

Regression models for systolic and diastolic blood pressure
UNDERUTILIZATION criterion

	Mean systolic blood pressure		Mean diastolic blood pressure	
Parameters	Beta Coefficients (S.E.)			
Constant	123.63 †	(5.00)	97.43 †	(2.52)
Underutilization (Criterion #53)	-0.91	(1.42)	-1.13	(.71)
Age	0.30 †	(.06)	-0.17 †	(.03)
# of Diagnostic categories	-0.50 †	(.31)	-0.28	(.16)
Compliance ratio	-0.34	(2.34)	-4.03 †	(1.17)
# of Antihypertensive drugs	2.37	(.60)	0.47	(.30)
Model Statistics				
R^2	0.09		0.07	
Adjusted R^2	0.08		0.06	
F (d.f.)	12.08 (609)		9.45 (608)	
p-value	0.00		0.00	

†$p<0.05$

Table 35

Regression models for systolic and diastolic blood pressure
OVERUTILIZATION criterion

	Mean systolic blood pressure		Mean diastolic blood pressure	
Parameters	Beta Coefficients (S.E.)			
Constant	124.12 †	(5.00)	97.02 †	(2.52)
Over-utilization (criterion #52)	2.12	(1.68)	0.31	(.84)
Age	0.30 †	(.06)	-0.17 †	(.03)
# of Diagnostic categories	-0.55	(.31)	-0.30	(.16)
Compliance ratio	-1.04	(2.42)	-3.89 †	(1.22)
# of Antihypertensive drugs	2.23 †	(.60)	0.40	(.30)
Model Statistics				
R^2	0.09		0.07	
Adjusted R^2	0.08		0.06	
F (d.f.)	12.35 (609)		8.93 (608)	
p-value	0.00		0.00	

†$p<0.05$

Table 36

Regression models for systolic and diastolic blood pressure
INDOMETHACIN AND DIURETICS drug-drug interaction criterion

	Mean systolic blood pressure		Mean diastolic blood pressure	
Parameters	Beta Coefficients (S.E.)			
Constant	106.89 †	(7.20)	95.00 †	(3.66)
Indomethacin and diuretics drug-drug interaction (criterion #36)	-10.84 †	(5.40)	0.51	(2.71)
Age	0.47 †	(.08)	-0.14 †	(.04)
# of Diagnostic categories	-0.03	(.41)	-0.26	(.21)
Compliance ratio	-1.04	(2.72)	-5.25 †	(1.36)
# of Antihypertensive drugs	3.10 †	(.80)	0.63	(.40)
Model Statistics				
R^2	0.14		0.07	
Adjusted R^2	0.13		0.05	
F (d.f.)	11.70 (360)		5.19 (359)	
p-value	0.00		0.00	

†$p<0.05$

Table 37

Regression models for systolic and diastolic blood pressure
CHOLESYTRAMINE/COLESTIPOL AND POTASSIUM WASTING DIURETICS
drug-drug interaction criterion

	Mean systolic blood pressure		Mean diastolic blood pressure	
Parameters	Beta Coefficients (S.E.)			
Constant	108.31 †	(7.20)	94.65 †	(3.64)
Cholesytramine/colestipol and potassium wasting diuretics drug-drug interaction (criterion #31)	-8.81	(8.81)	-0.72	(4.39)
Age	0.45 †	(.08)	-0.14 †	(0.4)
# of Diagnostic categories	-0.12	(.41)	-0.24	(.21)
Compliance ratio	-1.16	(2.74)	-5.20 †	(1.37)
# of Antihypertensive drugs	3.23 †	(.81)	0.64	(.40)
Model Statistics				
R^2	0.13		0.07	
Adjusted R^2	0.12		0.05	
F (d.f.)	11.05 (358)		5.12 (357)	
p-value	0.00		0.00	

†$p<0.05$

Table 38

Regression models for systolic and diastolic blood pressure
TRICYCLIC ANTIDEPRESSANTS AND ADRENERGIC AGENTS drug-drug interaction criterion

	Mean systolic blood pressure		Mean diastolic blood pressure	
Parameters	Beta Coefficients (S.E.)			
Constant	113.27 †	(18.71)	82.10 †	(9.90)
Tricyclic antidepressants and adrenergic agents drug-drug interaction (criterion #51)	15.09	(11.36)	3.56	(6.01)
Age	0.44 †	(.21)	-0.06	(.11)
# of Diagnostic categories	-1.13	(1.09)	0.03	(.58)
Compliance ratio	10.17	(9.19)	5.27	(4.87)
# of antihypertensive drugs	3.20 †	(1.35)	0.63	(.72)
Model Statistics				
R^2	0.24		0.05	
Adjusted R^2	0.18		-0.02	
F, *df*	4.15 (65)		0.73 (65)	
p-value	0.00		0.60	

†$p<0.05$